Molecular and Phylogenetic Analysis of a *Bacillus thuringiensis* Genetic Locus.

by

Alison Mowbray

December 1998

Gonville & Caius College, Cambridge

A Dissertation submitted for the degree of Doctor of Philosophy
in the University of Cambridge

Success is the ability to go from failure to failure without losing your enthusiasm.

Winston Churchill

Preface

This dissertation and experimentation herein are the result of efforts during the period from October 1993 to December 1998 in the Department of Biochemistry, University of Cambridge.

Certain Figures and Tables in this Thesis are reproduced from the work of other authors who are cited in the text.

This Thesis is my own unaided work and includes nothing which is the outcome of work done in collaboration, unless specific acknowledgement is made in the text. It has not been submitted to any other University.

Alison Mowbray

Contents

Acknowledgements

To my parents whose love, support, practical help and food parcels continue to enable me to pursue my far fetched and impoverishing ambitions.

To my supervisor Dr David Ellar for his help and patience and for proof reading this entire Thesis at least twice.

To the 'Lab Lasses': Sarah, Vish, Eileen, Cathy and Androulla for helping to make the last couple of years of this PhD not only bearable but almost enjoyable.

To all of 'Skylab' for their help and good humour especially Mike Dunn, Steve Gash, Trevor Sawyer, Graham Armstrong and David Hinks.

To some amazing friends: Liz, Morag, Chris and Julian for feeding, clothing and putting up with me in Cambridge and Jo, Kate and Stuart for feeding, clothing and putting me up in Putney.

To Richard Summers, Kim Rosewall and Chris Green for excellent photographic work.

To the Medical Research Council and everyone who ever bought a flapjack for their financial support.

And finally to Jemima P. for my sanity and always restoring my *joie de vie*.

Thankyou.

Summary

The diptericidal *Bacillus thuringiensis* (Bt) ssp. *fukuokaensis* strains 84-I and 17A were investigated for the presence of novel Cry proteins. N-terminal amino acid, immunological and PCR analysis indicated that both strains contain a novel set of δ-endotoxins. N-terminal amino acid sequence analysis indicated that the larger proteins from each strain (90 and 72-kDa of 84-I and 70 and 65-kDa of 17A) were related to the Cry proteins of Bt ssp. *israelensis*(Bti). Immunoblotting experiments confirmed that Cry10A-type proteins were present in both strains although subsequent PCR did not give a positive reaction for either strain using *cry10A* specific primers indicating that the Cry10-types were indeed novel.

To further investigate the 65-kDa protein of 17A, the gene encoding it was cloned from a size-enriched plasmid DNA library. Unsuccessful attempts were also made to clone the 90-kDa protein of 84-I. Sequence alignments of the deduced protein product of the 17A gene (*am1*) showed it to represent the second identification of a natural C-terminal truncate of a Cry4-type protein, the first being Cry10A. The missing C-terminal region of AM1 appears to be encoded as a complete Orf (*am2*) immediately downstream of the first protein gene.

When DNA containing both the *am1* and *am2* genes was subcloned into the pSVP27A expression vector high levels of expression of both proteins were observed in acrystalliferous Bt. The protein was deposited in inclusion bodies which were found to be toxic to *Dacus oleae*.

Extensive phylogenetic analysis was carried out to determine the relationship between, and possible evolutionary origins of, AM1, the Cry proteins of Bti and two further Cry10A-type δ-endotoxins (Cry19A from Bt ssp. *jegathesan* and Cry20A from 84-I) identified in other laboratories during the course of this project. Based on the amino acid sequence alignment, all seven proteins appear to have evolved from a common ancestor to form three distinct groups which mirror the structural organisation of the genes. Based on these groupings and a previous hypothesis of Dervyn *et al.* (1995), a hypothesis was proposed as to the evolution of the 130-kDa Cry4-type proteins from a 70-kDa Cry2-type ancestor.

The above hypothesis is based on the assumption that transfer of δ-endotoxin genes between subspecies has occurred at some point in evolutionary history. Evidence for this transfer was found when the genetic context of the *am1* gene was investigated. Two novel insertion sequences (T1) and (T2) were identified with sequence similarity to IS*240A* from Bti and an insertion sequence associated with the *Orf1* gene of 84-I. The identification of a further incomplete reading frame with similarity to integrase/recombinase proteins involved in Class II transposition raises the possibility that T1 and T2 form part of a novel Class II transposon.

A novel α/β-type small, acid soluble protein (SASP) gene was also discovered. This gene, which may be plasmid encoded, showed considerable sequence similarity to α/β-type SASP from *Bacillus megaterium*. The discovery of this gene raises new questions about taxonomic relations between the *Bacilli*.

Abbreviations

A. aegypti	*Aedes aegypti*
ATP	adenosine triphosphate
BBMV	Brush border membrane vesicles
bp	base-pair(s)
BSA	bovine serum albumin
Bt	*Bacillus thuringiensis*
Bti	*Bt* subspecies *israelensis*
CCY	casein casein yeast extract broth
Da	Dalton(s)
dATP	deoxyadenosine triphosphate
dH_2O	deionized water
DIG	digoxygenin
DNA	deoxyribonucleic acid
dNTPs	deoxynucleoside triphosphates
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylene diamine tetra acetic acid sodium salt
IPTG	isopropyl-β-D-thiogalactopyranoside
IS	Insertion Sequence
kb	kilobase-pair(s)
kDa	kiloDalton(s)
LB	Luria-Bertani broth
LC_{50}	concentration at which 50% mortality occurs
M	Molar
µg	microgram(s)
µg	milligram(s)
µl	microlitre(s)
Mda	megadalton
ml	millilitre(s)
µm	micrometre(s)
mM	milliMolar
mRNA	messenger ribonuleic acid
M. sexta	*Manduca sexta*
nm	nanometre(s)
OD_{600}	optical density at a wavelength of 600 nanometres
oligo	oligonucleotide

orf	open reading frame
PMSF	phenyl methyl sulphonyl fluoride
PWYE	peptone water yeast extract broth
RNA	ribonucleic acid
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SDS/PAGE	SDS/polyacrylamide gele electrophoresis
Subsp./ssp.	subspecies
TAE	Tris/acetate/EDTA buffer
TBE	Tris/borate/EDTA buffer
TBS	Tris buffered saline
TE	Tris EDTA buffer
Tn	Transposon
X-Gal	5-bromo-4-chloro-3-indolyl-β-galactopyranoside
x *g*	units of gravitational force

One

Chapter 1

Introduction

1.1 *Bacillus thuringiensis*:- Background Information

With a market share of over 95% (Kvistgaard, 1994) *Bacillus thuringiensis* (Bt) is the most commonly used microbial insecticide in the world. A broad range of bacteria, viruses and fungi are pathogenic to insects (Aronson *et al.*, 1986). However, the formation of Bt insecticidal proteins into easily harvestable cytoplasmic crystals during sporulation and the toxicity of these proteins to a wide range of major insect crop pests and disease vectors (Höfte & Whiteley, 1989) has ensured that the popularity of Bt in microbial insect control is as yet unchallenged.

First isolated in 1901 in Japan by Ishiwata (Dulmage & Aizawa, 1982) as the causative agent of the "sotto" disease in silkworms (*Bombyx mori*), the name *Bacillus thuringiensis* was coined by Berliner (1915) to describe the pathogenic bacteria found in the larvae of the Mediterranean fruit fly (*Ephestia kuhniella*). Bt is a Gram-positive bacterium, which produces proteinaceous inclusions during sporulation.

According to many studies (Ash *et al.*, 1991; Nakamura, 1994) the identification of these distinctively shaped, cytoplasmic crystals by phase contrast microscopy provides the only means of distinguishing *Bacillus thuringiensis* from the very closely related *Bacillus cereus*. This character is generally plasmid encoded and transmissible to *B. cereus* by conjugation (Damgaard, 1995; Gonzales *et al.*, 1982) and there is in fact considerable evidence that *B. thuringiensis* and *B. cereus* should be considered a single species. Some strains previously believed to have been *B. cereus* were found to hybridise to Bt crystal-specific probes (Carlson *et al.*, 1994) and more recently, specific DNA probes based on the variable regions of 16S rRNA used to discriminate between the two species (te Giffel *et al.*, 1997) yielded as many false positives and negatives as accurate identifications (te Giffel *et al.*, 1997). However, despite mounting evidence to the contrary, current taxonomy still places Bt and *B. cereus* in separate species (Damgaard, 1995).

There is some debate as to the natural habitat of Bt. Populations are so routinely isolated from dead and dying insects that there can be little doubt as to its existence as an insect pathogen; the ability of Bt to multiply in the insect haemocoel and to provoke septicemia has long been recognised (Stephens, 1952; Heierson *et al.*, 1986). This virulence is thought to be due not only to the presence of insecticidal crystal proteins but also to a large number of extracellular compounds that are discussed later in this Chapter. It is however

questionable whether Bt is able to colonise other habitats outside the insect host. Many studies indicate that although Bt may be isolated from a variety of natural habitats such as soil, it does not survive or grow well in these environments (Tapp & Stotzky, 1995). One study in particular demonstrated that germination of Bt ssp. *kurstaki* occurred in dead *Pieris brassicae* larvae but not in soil or leaf samples (Pederson *et al.*, 1995). This is contradicted however by recent ecological studies that have shown that Bt occurs naturally at high frequencies on the phylloplanes of herbaceous plants and arbores (Ohba, 1996; Damgaard *et al.*, 1994). Such studies may suggest that the Bt populations commonly found associated with plant-derived materials, such as stored product environments (Burges & Hurst, 1977; DeLucca *et al.*, 1982), animal feed mills (Meadows *et al.*, 1992) and rice mill factories (Chanpaisang *et al.*, 1994) are of plant origin. A greater understanding of the true ecological roles of Bt is important if the effects of releasing Bt as bio-insecticides are to be properly understood. Such information should also assist in isolating novel Bt strains containing further useful insecticidal protein genes.

The Bt cytoplasmic crystals are composed of one or more types of insoluble polypeptides called delta (δ)-endotoxins which are the primary cause of Bt insecticidal activity. The crystals develop during sporulation and upon completion of sporulation the bacteria lyse releasing the spore and crystal into the environment where they are available for ingestion by insect larvae. Different isolates of Bt can produce different crystal proteins, each of which may be toxic to a different insect species (Höfte & Whiteley, 1989; Lambert & Peferoen, 1992). Intensive screening programmes have identified strains showing activity towards a wide range of important agricultural and health related insect pests including Lepidoptera (butterflies and moths), Diptera (mosquitoes and flies), Coleoptera (beetles), Hymenoptera (ants) and Acaria (mites) (Feitelson, 1993); as well as other invertebrates such as the nematode parasites of mammals and plants (Edwards *et al.*, 1990; Bone *et al.*, 1988; Bone, 1989), the trematode (animal parasitic liver flukes), the acari (mites) and some protozoan pathogens (Feitelson, 1993).

The increasingly large number of Bt isolates has required the formation of a reliable and reproducible classification scheme. The diversity of flagellar H-antigen agglutination reactions has provided the best such scheme (de Barjac & Bonnefoi, 1968; de Barjac & Frachon, 1990) and to date, 55 different flagella serotypes and 8 nonflagellated biotypes have been catalogued (Lecadet *et al.*, 1994; Schnepf *et al.*, 1998).

Whilst this classification system works well for Bt subspecies it provides little predictive value as to the insect host range of individual isolates. Activity towards a target insect is defined by the number and type of δ-endotoxins that make up the crystalline inclusion. It is not surprising therefore that some isolates which share the same insect pathogenicity have different serotypes. Conversely, it may be true that two isolates which have the same serotype could display different toxicities. It is for this reason that δ-

endotoxins have their own classification system (Crickmore *et al.*, 1998) and this is discussed in section 1.4.

1.2 Commercial Application

Insects are responsible for the loss of 20% of all major crops before harvest and the transmission of many of the most crippling human diseases including yellow fever, malaria and river blindness. In the battle against these pests, the overuse and misuse of synthetic insecticides, as well as poor understanding of their mechanism of action, has caused serious problems. These include harm to non-target organisms, environmental pollution and insect resistance. Escalating environmental concerns over the use of traditional chemicals like DDT and parathion, of which generations of humans have been almost totally dependent, has spurred the search for 'greener' biological alternatives.

Despite the discovery of insecticidal activity at the beginning of the century, it was not until the 1950's that development of Bt formulations to control damaging caterpillars in agriculture and forestry (Rowe & Margaritis, 1987) provided the first real alternative to chemicals for the control of insect pests. Since then, nearly 200 products based on isolates of Bt have been registered to control insect pests including caterpillars, mosquitoes, black flies and beetles (Figure 1.1). Spore/crystal preparations are produced by culturing the *Bacillus thuringiensis* strain to sporulation in industrial fermenters. The mixture of spores and crystals is then centrifuged, dried and formulated as a powder or liquid spray. Such preparations from various strains have been marketed as insecticides for nearly 40 years (Van Frankenhuyzen, 1993; Dulmage, 1981).

The environmental advantages of Bt insecticides over their chemical counterparts are clear. First, despite the use of *Bacillus thuringiensis* for nearly 40 years there have been relatively few reported cases of field-resistance and these are restricted to larvae of a single species, the diamond back moth, *Plutella xylostella* (Kirsch & Schmutterer, 1988; Tabashnik *et al.*, 1990; Hama *et al.*, 1992; Shelton *et al.*, 1993). Second, unlike many chemical pesticides, Bt insecticides are completely non-toxic to mammals, birds, amphibian and reptiles yet have highly specific activity to individual or small groups of insects and invertebrate pests. This enables pests to be targeted without harming non-target vertebrates and invertebrates (Briggs, 1986). The high specificity is thought to be due to the steps involved in activation of the toxins and their binding to insect gut receptors which are discussed later in this Chapter. Because of their low toxicity to predators and parasites (Salama *et al.*, 1982), the Bt toxins do not give rise to an explosion of pest numbers after the initial decline following treatment. Nor, for the same reason, does the application of Bt

Agent	Active ingredient(s)	Crop	Year registered	No. of products	Target pest
Bacterium	*B. popilliae, B. lentimorbus*		1948	2	Japanese beetle larva
	Bt ssp. *kurstaki*		1961	127	Lepidopteran larva
	Bt ssp. *israelensis*		1981	26	Dipteran larva
	Bt ssp. Berliner		1984	1	Lepidopteran larva
	Bt ssp. *tenebrionis*		1988	6	Coleopteran larva
	Bt ssp. *kurstaki*EG2348		1989	4	Lepidopteran larva
	Bt ssp. *kurstaki* EG2424		1989	1	Lepidopteran larva
	Bt ssp. *kurstaki* EG2371		1990	3	Lepidopteran larva
	B. sphaericus		1991	1	Dipteran larva
	Bt ssp. *aizawai* GC-91		1992	2	Lepidopteran larva
	Bt ssp. *aizawai*		1992	2	Lepidopteran larva
	Bt ssp. *kurstaki* BMP123		1993	5	Lepidopteran larva
	Bt ssp. *kurstaki* EG7673		1995	2	Lepidopteran larva
	Bt ssp. *kurstaki* EG7673		1995	2	Colorado potato beetle
	Bt ssp. *kurstaki* EG7841		1996	1	Lepidopteran larva
	Bt ssp. *kurstaki* EG7826		1996	3	Lepidopteran larva
	Bt ssp. *kurstaki* M200		1996	1	Lepidopteran larva
Nonviable microbial pesticide	Bt ssp. *kurstaki* δ-endotoxin in killed *P. fluorescens*		1991	2	Lepidopteran larva
	Bt ssp. *san diego* δ-endotoxin in killed *P. fluorescens*		1991	1	Coleopteran larva
	Bt Cry1Ac and Cry1C δ-endotoxin in killed *P. fluorescens*		1995	1	Lepidopteran larva
	Bt ssp. *kurstaki* Cry1C δ-endotoxin in killed *P. fluorescens*		1996	1	Lepidopteran larva
Plant pesticide	Bt Cry3A δ-endotoxin	Potato	1995	1	Colorado potato beetle
	Bt Cry1Ab δ-endotoxin	Corn	1995	2	Lepidopteran larva
	Bt Cry1Ac δ-endotoxin	Cotton	1995	1	Lepidopteran larva
	Bt Cry1Ab δ-endotoxin	Corn	1996	2	Lepidopteran larva
	Bt ssp. *kurstaki* δ-endotoxin from HD-1 derived plasmid vector pZ01502	Corn	1996	2	Lepidopteran larva
	Bt ssp. *kurstaki* Cry1Ac δ-endotoxin	Corn	1997	1	Lepidopteran larva

Figure 1.1 Microbial pesticides registered by the U.S. Environmental Protection Agency as of 1997.
Taken from Schnepf *et al.*, 1998.

pesticides lead to the creation of secondary pests that had previously been contained by natural enemies (Salama *et al.*, 1982). The third environmental benefit is that, because the parasporal crystals are susceptible to UV degradation (Pusztai *et al.*, 1991; Liu *et al.*, 1993), wash off by rain and bio-degradation by micro-organisms, Bt insecticides have low persistence. Therefore, unlike their chemical counterparts they do not persist in the soil or in the food chain after they have served their purpose.

This transfer of emphasis to environmentally friendly pesticides has perhaps had most impact in the forests of the United States. Here Bt has become the major pesticide used against the Gypsy moth (Mackesky, 1989), by no means the only forest pest successfully controlled with Bt. Currently targeted pests include the spruce budworm (Canada), the nun moth (Poland), the Asian gypsy moth (United States, Canada and the Far East), the pine processional moth (Spain and France), and the European pine shoot moth (South America) (Bowen, 1996, unpublished).

Bearing all this in mind, it is perhaps surprising that in 1995, world-wide sales of Bt were projected at $90 million (Schnepf *et al.*, 1998), just 2% of the total global insecticidal market (Lambert & Peferoen, 1992). Despite having development costs estimated at 1/40 that of a comparable novel synthetic chemical pesticides (Becker & Margalit, 1993), Bt insecticides still appear to be having a hard time competing with an established, recognised chemical pesticide industry. This is due, in large part, to the fact that the very features of Bt δ-endotoxins that make them 'green' can actually reduce their commercial appeal. For instance the specificity of Bt toxins has to some extent limited their value as general pesticides. If a range of pests has to be eliminated, then a range of products must be used and knowledge of exactly which species are threatening which areas is required before application. This narrow spectrum of insecticidal activity is also a problem in that several insect pest species of global importance have little or no susceptibility to δ-endotoxin from existing Bt strains.

Low persistence is also a disadvantage as Bt products often have to be applied more frequently than chemical insecticides. Even the prodigious appeal of virtually resistance-free insecticides is tempered by the prediction that with continued use, Bt resistance is likely to evolve in insects other than the diamond back moth. Repeated selection in the laboratory with either sporulated cultures or purified toxin has produced resistant strains in other moths (Gould *et al.*, 1992; McGaughey, 1985; McGaughey & Johnson, 1994; Moar *et al.*, 1990; Stone *et al.*, 1989) and beetles (Rahardja & Whalon, 1995). However, despite these gloomy predictions, an ever increasing battery of Bt δ-endotoxins (many with novel toxicities) and significant advances in genetic engineering (some of which are described below) means that interest in Bt to control insect pests is still high.

The isolation of numerous new *Bacillus thuringiensis* strains is becoming routine in many laboratories, expanding the range of pests that can be targeted and providing

alternatives for the control of individual insect pests. It also offers a strategy for coping with problems of resistance as a mixture of different toxins with activities against the same target insect may often prove more effective than a single toxin in delaying the onset of resistance (Gill *et al.*, 1992; Tabashnik *et al.*, 1991; Van Rie *et al.*, 1990b).

Genetic engineering of δ-endotoxins has also helped remove obstacles to their effective use. To try and overcome the limiting narrow specificity of natural Bt isolates and increase the target range of a single preparation, several genes toxic to different insects can be introduced into single strains of bacteria (Crickmore *et al.*, 1990). This manipulation can be achieved in a variety of ways. A conjugation-like system has been used to transfer crystal encoding plasmids from one strain to another, resulting in transconjugant strains with improved insecticidal activity (Gawron-Burke & Baum, 1991). This process mirrors the natural process of gene transfer between bacteria and is therefore subject to the least restrictions for use in the field.

Many δ-endotoxin genes are not readily transmissible via conjugation but most strains can now easily be transformed with plasmid DNA using electroporation procedures (Bone & Ellar, 1989; Mahillon *et al.*, 1989). Shuttle vectors (Lereclus *et al.*, 1989; Baum *et al.*, 1990; Crickmore *et al.*, 1990) have further facilitated the expression of several cloned δ-endotoxin genes in Bt transformants. The cloned δ-endotoxin genes may be maintained in Bt on recombinant plasmids or else integrated into resident plasmids or the chromosome by homologous recombination (Adams *et al.*, 1994; Baum *et al.*, 1990; Kalman *et al.*, 1995; Lereclus *et al.*, 1992). With these approaches, combinations of δ-endotoxin proteins can be designed rather than discovered and crystal protein production can be augmented by selecting host backgrounds suitable for large-scale fermentation. In addition, δ-endotoxins engineered for improved insecticidal activity, fermentation yield, or stability can be used as active ingredients, further expanding the possibilities for improving Bt based bio-pesticides.

Despite such advances, biological control (Bt or otherwise) is too often just trying to catch up with a well understood, established chemical industry. Better the devil you know than the devil you don't. The reality is, that wherever chemical and biological insecticides are in direct competition, biological control will too often remain a limited experiment in alternative technology. In order to succeed, biological control must open up avenues of pest management that are not available to their chemical counterparts and establish their own niche markets. Genetic engineering has created such opportunities.

One particular example is the ineffectual nature of all surface applied insecticides (synthetic and biological) against pests with cryptic feeding habits i.e. those that feed internally or on the roots of the plant. Genetic engineering of biological insecticides has opened up two main avenues in the battle against these pests. Avenues in which chemical insecticides cannot compete.

Firstly, the crystal toxin genes can be introduced into a host that can propagate itself at the site of feeding and continue to produce crystal protein. The rhizosphere and rhizoplane are heavily infested with saprophytic micro-organisms that are well adapted to compete under low nutrient levels and may well prove suitable hosts. Such a relationship not only improves the persistence of the Cry proteins but expands their delivery range to include sucking, boring and root dwelling insects and nematodes. For example the *cry1Ac* gene has been introduced into the endophytic bacterium *Clavibacter xyli* on an integrative plasmid (Lampel *et al.*, 1994), and the resulting recombinant strain used to inoculate corn to control European corn borer infestations (Tomasino *et al.*, 1995). *Rhizobium leguminosarum* biovars have also been used to deliver the *cry3A* gene into the nodules of pea (*Pisum sativum*) which then caused mortality among feeding pea weevil (*Sitona ssp.*) larvae (Skøt *et al.*, 1990).

Alternative delivery systems have also been sought to increase the persistence of the mosquitocidally active Bt ssp. *israelensis* (Bti) toxins in the aquatic larval feeding zone. In the laboratory, mosquito larvae fed on *Tetrahymena pyriformis* loaded with Bti died three times faster in the case of *Aedes aegypti* (Manasherob *et al.*, 1994) and eight times faster in the case of *Aedes stephensi* (Manasherob *et al.*, 1996) than when fed on the same concentrations of Bti alone. This is probably as a consequence of concentrating large quantities of Bti spores and crystals in *T. pyriformis* cells and floating them on the water surface in the larval feeding zone. Without such measures the toxin can quickly sediment and under field conditions would become lost in the mud. Transgenic *Cyanobacteria* (Angsuthanasombat & Panyim, 1989; Soltesrak *et al.*, 1995) and *Caulobacter crescentus* (Thanabalu *et al.*, 1992; Yap & Porter, 1993) expressing mosquitocidal δ-endotoxin genes have also been developed to improve the persistence of the toxin in the upper layers of water where the target larvae are found.

The second avenue opened up by genetic engineering is the expression of Bt toxin genes in the plant itself. This provides the plant with self-protection against pests, therefore overcoming the problems associated with other toxin delivery systems. Several plants have been successfully transformed with Bt δ-endotoxin genes (which do need to be modified before significant expression is seen) starting with tobacco (Barton *et al.*, 1987; Vaeck *et al.*, 1987) and currently including more than fifty types of plant (see Peferoen, 1997 for a review). Potato, cotton and corn containing modified *cry* genes have been sold to growers since 1996.

Transgenic plants and micro-organisms offer a means of overcoming the problems of delivering Bt toxins to a target insects. However, it is likely to increase the risk of resistance development due to prolonged presence of the δ-endotoxin throughout the season. Detailed knowledge of the Bt genetics, structure and mechanism of action are therefore of crucial importance in identifying and combating insect resistance mechanisms. This is necessary to

secure the long term future of Bt bio-pesticides as it seems likely that a major opportunity for the development of Bt bio-pesticides lies in the control of insect pests showing resistance to agrochemicals.

Concerns have also been raised about the ecological and health impact of disseminating genetically engineered micro-organisms, particularly those containing antibiotic resistance genes (Fox, 1995), and attempts are being made to limit the release of foreign DNA elements. One such example is the development of a 'homologous' expression vector system combining a Bt plasmid replicon with an indigenous site-specific recombination system (Baum *et al.*, 1996). Ancillary or foreign DNA can be selectively removed from the recombinant bacterium after introducing the δ-endotoxin-encoding plasmid vector, thus eliminating the necessity of maintaining copies of an antibiotic resistance gene or any foreign DNA elements within the recombinant strain. Using this system a number of Bt strains were constructed (Baum *et al.*, 1996) which harboured unique combinations of crystal proteins not as yet found in nature, some of which exhibited improved insecticidal properties and/or improved crystal protein yield.

In addition, the impact of gene flow to wild relatives needs to be assessed. Preliminary experiments show that cross-hybridisation among members of the Brassicaceae family and *Brassicae napus* carrying a Bt gene may be possible under certain conditions (Stewart, 1997). This may imply that transgenic *B. napus* may transfer its insecticidal Bt gene to wild relatives (Stewart, 1997).

1.3 Pathogenic Determinants

While the δ-endotoxins are the most important factors in Bt insect pathogenesis, other minor toxic factors are produced during the growth cycle of the organism. These include α and β-exotoxins, enterotoxin, the 'louse' factor with activity against certain mammalian biting lice (Gingrich *et al.*, 1974) and various exoenzymes and immune inhibitors.

The heat-labile α–protein has been demonstrated to be active against mice and several lepidopteran insects (Krieg, 1971). The heat-stable β-exotoxins are nucleotide analogues produced during vegetative growth (Sebesta *et al.*, 1981) and their chemical structures differ from strain to strain (Levison *et al.*, 1990). Their toxicity is thought to arise from a common mechanism of inhibition of DNA-directed RNA polymerases (Sebesta & Sternbach, 1970) and this would explain their wide toxicity including vertebrates and invertebrates. Because of these properties, commercial preparations of Bt-based products marketed in Western Europe and the USA are not permitted to contain the β-exotoxin.

Recently, a range of *Bacillus thuringiensis* strains were found to produce the diarrhoeal enterotoxin that is routinely isolated from the very closely related food poisoning

organism *Bacillus cereus* (Damgaard, 1995). Nine strains from different commercial based Bt insecticides and a tenth isolated from a fatal case of bovine mastitis caused by Bt (the only reported case of an animal death from Bt infection (Gordon, 1977)) were tested and although the titres were low, all were found to produce the enterotoxin. Only a few incidents linking Bt to human infection have ever been described (Kramer & Gilbert, 1989; Jackson *et al.*, 1995) but, the separation of *B. thuringiensis* from *B. cereus* requires techniques which are not used in identifying cases of food poisoning by *B. cereus* suggesting that the role of Bt in food poisoning may have been underestimated. Commercial products of Bt-based insecticides, spores and δ-endotoxin are however not toxic to mammals because the pathogenicity in terms of food poisoning by diarrhoeal enterotoxin, is caused by ingestion of vegetative cells.

A new family of insecticidal proteins (vegetative insecticidal proteins (Vips)) produced by Bt during its vegetative stages of growth has recently been discovered (Warren *et al.*, 1994) that bear no similarity to δ-endotoxins. The *vip3A* gene encodes an 88-kDa protein that is secreted into the supernatant fluid by Bt cultures (Estruch *et al.*, 1996) and displays insecticidal activity against a wide range of lepidopteran insects. It appears that the midgut epithelium cells of susceptible insects are the primary target for the Vip3A insecticidal protein and that their subsequent lysis is the primary mechanism of death (Yu *et al.*, 1997). The histopathology associated with the ingestion of Vip3A resembles the well-documented series of events that follow the ingestion of δ-endotoxin by susceptible insects and this is discussed in Section 1.6.

Chitinase (Luthy, 1980), phospholipase C (Taguchi *et al.*, 1980) and haemolysin (Pendleton *et al.*, 1973) are all exoenzymes produced by vegetative Bt that may also play an important role in its overall pathogenicity. Other important enzymes produced by Bt may be the two immune inhibitors, InA and InB that might enable Bt to invade the insect immune system. InA is a metallopeptidase common in Bt subspecies but virtually absent in other bacilli (Lovgren *et al.*, 1990). This enzyme specifically inactivates the antibacterial protein attacins and cecropins in the haemolymph. In contrast InB appears to be a non-protein substance which can inhibit the insects' response against *B. cereus* (Edlund *et al.*, 1976).

1.4 The δ-Endotoxins

The major toxin produced by *Bacillus thuringiensis* is the proteinaceous δ-endotoxin which forms a cytoplasmic crystalline inclusion concomitantly with sporulation (Bechtel & Bulla, 1976). δ-endotoxins have an insecticidal effect on a number of insects and have also been reported to suppress the growth of mammalian tumour cells and exhibit antibiotic activity in relation to certain micro-organisms (Yudina *et al.*, 1988; Yudina *et al.*, 1996;

Yudina & Burtseva, 1997). A single crystal may contain one million protein subunits held together by interchain disulphide bonds (Du *et al.*, 1994) and can account for as much as 20 to 30% of the cell's dry weight (Ellar, 1990).

The process of sporulation and crystal formation can be divided into seven major stages (Ryter, 1965) which are characterised by morphological changes. A diagram representing this process is shown in Figure 1.2.

Synthesis of the δ-endotoxin is initiated during stages II to III (Lecadet & Dedonder, 1971; Bechtel & Bulla, 1976), and is greatest during stage IV, although final maturation of the crystal may continue to the last stages of sporulation (Ribier & Lecadet, 1973). The δ-endotoxin accumulates alongside the spore in the mother cell until breakdown of the cell wall releases the mature spore and δ-endotoxin into the environment.

As well as its existence as discrete crystals, δ-endotoxin has been reported to be attached to the surface of the spore (Du & Nickerson, 1996) and in at least two strains, the crystal remains attached to the spore after lysis (Debro *et al.*, 1986; Lopez-Meza & Ibarra, 1996). These discoveries add weight to the suggestion that the evolutionary origins of the crystal may arise from the over-production of spore coat proteins (Stahly *et al.*, 1978).

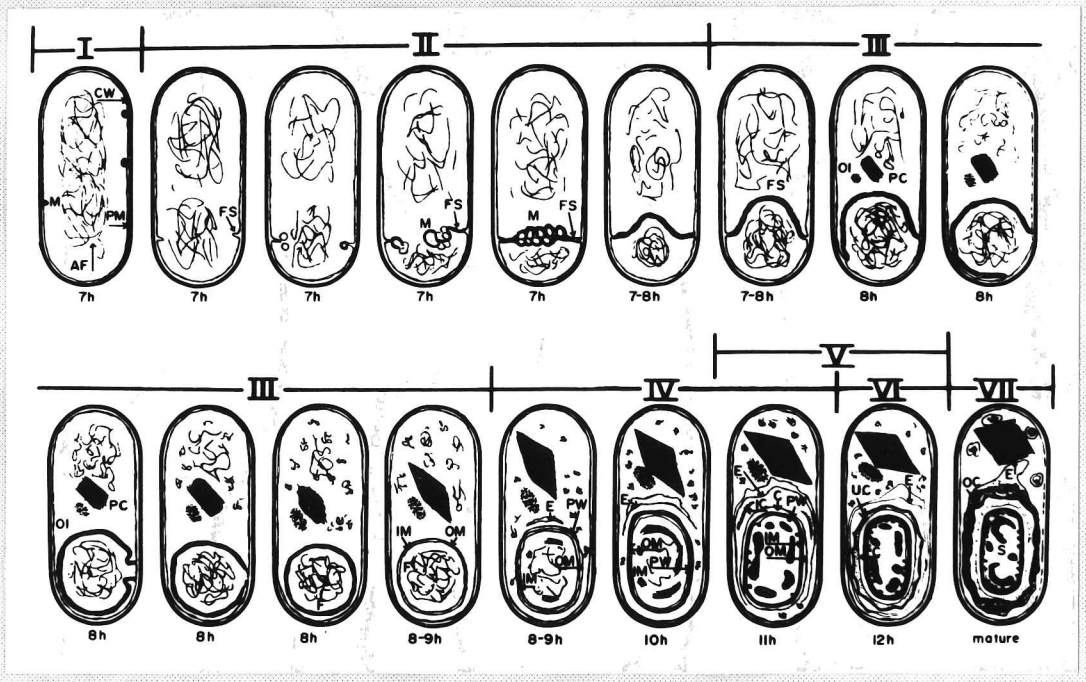


Figure 1.2 Diagrammatic scheme of sporulation in Bt.
M=mesosome; CW=cell wall; PM=plasma membrane; AF=axial filament; FS=forespore; OI=ovoid inclusion; PC=parasporal crystal; F=forespore; IM=inner membrane; OM=outer membrane; PW=primordial cell wall; E=exosporium; LC=lamellar spore; OC=outer spore coat; C=cortex; UC=undercoat; S=mature spore in an unlysed sporangium.
Reproduced from Bechtel and Bulla (1976). (photo no. 120A2).

Given that sporulation is a response to starvation, the synthesis of the δ-endotoxin during this time suggests that it is important for spore survival and germination. It may also confer a selective advantage for this organism over other bacteria. A suggestion as to the nature of this advantage has been proposed by Ellar *et al.* (1990). Following ingestion by an insect, the crystal δ-endotoxin is solubilised in the alkaline midgut and activated by host proteases. The cells lining the midgut are lysed by the active δ-endotoxin, causing a reduction in midgut pH, cessation of larval feeding and the equilibration of blood and gut contents. If a spore is ingested together with the δ-endotoxin crystal, the rich nutrient environment after cytolysis may enable Bt spores to initiate a new vegetative colony in an environment inimical to the germination of other spores. Other workers have suggested that the presence of crystal protein on the surface of the spore may cause the spore to be activated by the alkaline conditions in the midgut environment and then triggered for germination by binding to gut receptors (Du & Nickerson, 1996). Such accelerated spore germination could then provide invasive vegetative cells ready to participate in pathogenesis as soon as the gut pH drops sufficiently. Such a mechanism to increase the spore germination rate could be advantageous because of the comparatively short transit time for food passing through the larval gut.

The crystalline δ-endotoxins produced by Bt subspecies have a variety of different morphologies (Figure 1.3). These shapes are often characteristic of their insecticidal activity and thus their δ-endotoxin composition (Carlton *et al.*, 1990). Lepidopteran-specific toxins tend to form bipyramidal inclusions, dual-specific (lepidopteran and dipteran) form composite crystals and coleopteran-specific toxins have a flat-rhomboidal shape. Bt ssp. *israelensis* crystalline inclusions are irregular and consist of at least three separate structures surrounded by a unique envelope structure (Ibarra & Federici, 1986). In other subspecies, the δ-endotoxins tend to combine to form one single inclusion (Held *et al.*, 1990).

It has been found that cloned δ-endotoxin genes are often visibly indistinguishable from the crystals of the strain from which they were cloned. However, the biochemical properties, such as solubility, of these cloned toxins can be altered (Aronson *et al.*, 1991; Angsuthanasombat *et al.*, 1992) perhaps indicating that other factors are involved in the formation of some crystalline inclusions. The 20-kDa protein from Bt ssp. *israelensis* is one such factor (Adams *et al.*, 1989) and is discussed in Section 1.5. SDS/PAGE analysis of the crystalline inclusions reveal that generally, they are composed of one or more polypeptides with molecular masses ranging from 25 to 140-kDa (Figure 1.4).

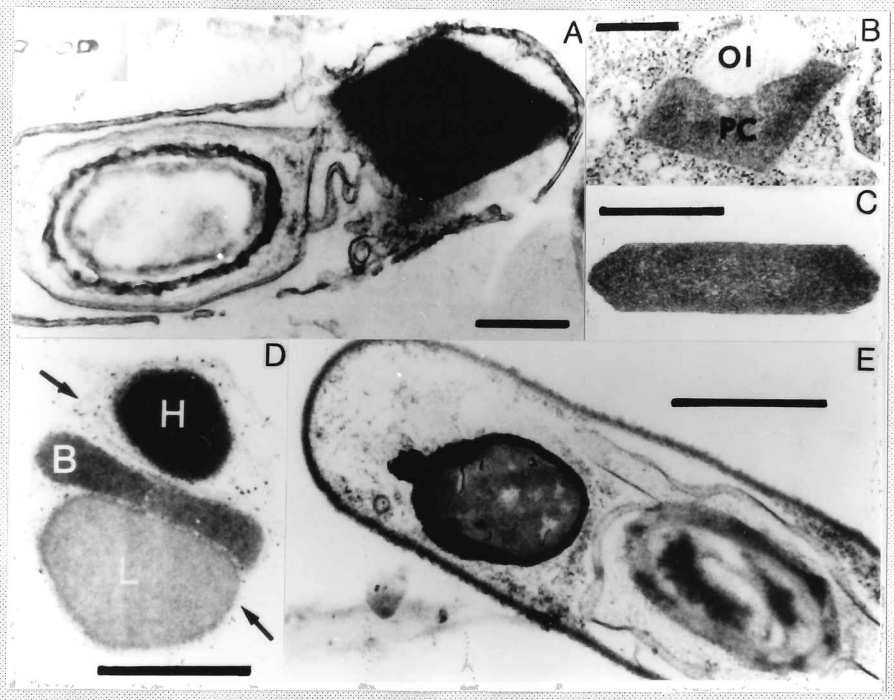


Figure 1.3 Morphology of Bt δ-endotoxins.
(A) Bipyramidal crystal of the lepidopteran-specific Bt ssp. *kurstaki* HD1 (Johnson *et al.*, 1980).
(B) Parasporal inclusion of Bt ssp. *kurstaki* HD1 showing the 63-kDa ovoid inclusion (OI) as well as the bipyramidal crystal (PC) (Lüthy *et al.*, 1982).
(C) Isolated crystal from the coleopteran-active Bt ssp. *san diego* (Herrnstadt *et al.*, 1986).
(D) Isolated crystal from Bt ssp. *israelensis* (Ibarra & Federici, 1986). L, large inclusion of low electron density; B, bar shaped body; H, inclusion of high electron density. Arrows show the unique envelope surrounding the crystal.
(E) Crystal of the dipteran active Bt ssp. *kyushuensis* (Held *et al.*, 1990).
All scale bars represent 0.5 μm. Photograph courtesy of P. Koni (photo no. 239C5).

Polyclonal antisera raised against purified crystals have shown the serological diversity of these proteins and led to the firm establishment of the δ-endotoxins as a heterogeneous sized family of proteins, each with a unique host range (Dulmage, 1981; Yamamoto & McLaughlin, 1981). The advent of molecular biology revealed detailed information regarding the amino acid sequence of these proteins and in 1989, with at least 42 crystal genes cloned and sequenced, a nomenclature and classification scheme was proposed based on sequence similarity and insecticidal activity spectrum of the different proteins (Höfte & Whiteley, 1989). The CryI proteins were toxic to lepidopterans; CryII to both lepidopterans and dipterans; CryIII to coleopterans and CryIV to dipterans alone. Cyt proteins were a class of much smaller proteins that were both diptericidal and broadly cytolytic *in vitro* and displayed no sequence similarity to the Cry proteins.

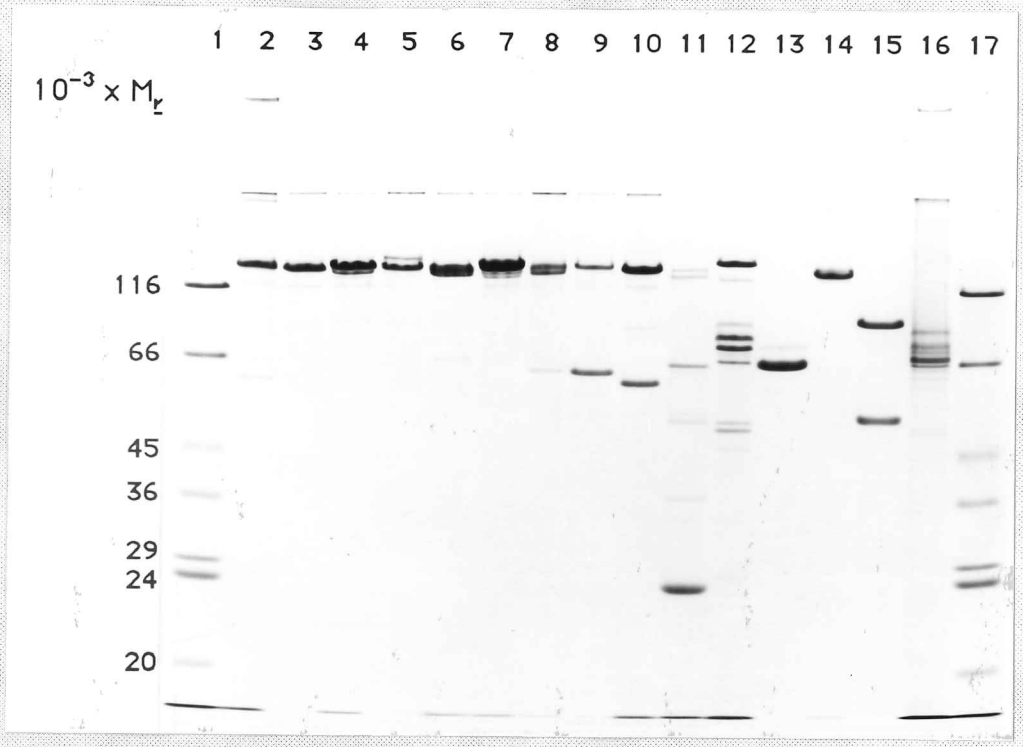


Figure 1.4 Coomassie blue stained 13% SDS/PAGE gel of polypeptides from various Bt subspecies. Photo courtesy of Dr. J. Carroll (photo no. 919B5).

Lane	Pathotype*	Lane	Pathotype*
1 Molecular weight standards		10 Bt ssp. *galleriae* SN916	L/D
2 Bt ssp. *sotto* SN913	L	11 Bt ssp. *israelensis*	D
3 Bt ssp. *kurstaki* HD73	L	12 Bt ssp. *darmstadiensis* 73-E-10-2	D
4 Bt ssp. *aizawai* HD229	L	13 Bt ssp. *tenebrionis*	C
5 Bt ssp. *entomocidus* 60/1	L	14 Bt ssp. *dakota*	NT
6 Bt ssp. *aizawai* IC1	L/D	15 Bt isolate DJE 783/2	NT
7 Bt ssp. *morrisoni* HD12	L/D	16 Bt ssp. *ostriniae*	NT
8 Bt ssp. *kurstaki* HD1	L/D	17 Molecular weight standards	
9 Bt ssp. *thuringiensis* HD770	L/D		

* Specified pathotypes; L, lepidoptera; D, diptera; C, coleoptera; NT, non-toxic to date.

While useful, this classification scheme created problems in assigning nomenclature to the ever increasing number of new δ-endotoxins. The main difficulty arose from the mixed criteria of sequence similarity and insecticidal activity spectrum used to construct the classification. For example, CryI toxins were supposed to be specific to Lepidoptera. However, CryIAb from Bt ssp. *aizawai* IC1 (Haider & Ellar, 1987b) and CryIC (Smith *et al.*, 1996) were found to exhibit dual activity against Lepidoptera and Diptera. CryIB has also been shown to exhibit toxicity to three orders of insects (Coleoptera, Lepidoptera and Diptera) (Zhong, 1996). Other toxin genes were also cloned displaying no sequence

similarity to the *cry* or *cyt* genes (Brown & Whiteley, 1992). For these reasons a new nomenclature was devised based only on amino acid similarity between the full length toxin sequences (Crickmore *et al.*, 1998).

This new nomenclature has recently been published (Crickmore *et al.*, 1998) and assigns names (replacing Roman numerals by Arabic numerals in the primary rank e.g. Cry1Aa) to crystal proteins from Bt that exhibit some experimentally verifiable toxic effect to a target organism, or, any protein that has obvious sequence similarity to a known Cry protein. The proteins are named according to their degree of evolutionary divergence as estimated by phylogenetic tree algorithms.

The change from a function-based to a sequence-based nomenclature allows closely related toxins to be ranked together. Detailed information about these relationships can then be conveyed by appending to the root a series of numbers and letters indicating degrees of phylogenetic divergence (Figure 1.5). This not only removes the necessity for researchers to bioassay each new protein against a growing series of organisms before assigning it a name, but can in fact indicate potential activities since proteins with the same primary rank often affect the same order of insect. Those with different secondary and tertiary ranks may have altered potency and targeting within an order that may be due to the accumulation of point mutations, but often appear to have resulted from ancestral recombination events between genes differing at a lower rank level (Bravo, 1997). The quaternary rank was established to differentiate between genes coding for known toxins that differ only slightly, either because of a few mutational changes or an imprecision in sequencing (Crickmore *et al.*, 1998).

Regardless of what the δ-endotoxins were called when they were first published the new nomenclature will be used during the rest of this Thesis. Where reference is made to the old nomenclature this will be indicated in the text. The complete scheme is available on the World Wide Web (http://www.susx.ac.uk/Home/Neil Crickmore/Bt/index.html).

1.5 Location and Expression of δ-Endotoxin Genes

Location

Crystal protein synthesis can be lost spontaneously (Gonzalez *et al.*, 1981) through prolonged storage (Gordon *et al.*, 1973), medium shift (Rajalakashmi & Shethna, 1977), or by growth at elevated temperature (42°C) (Carlton & Gonzalez, 1985), indicating that the δ-endotoxin genes are located on plasmids (Gonzalez *et al.*, 1981). Direct evidence of this was obtained when the *cry1Aa* gene was cloned from Bt ssp. *kurstaki* HD1 Dipel plasmid DNA (Schnepf & Whiteley, 1981). For several well-characterised strains, the protoxin genes

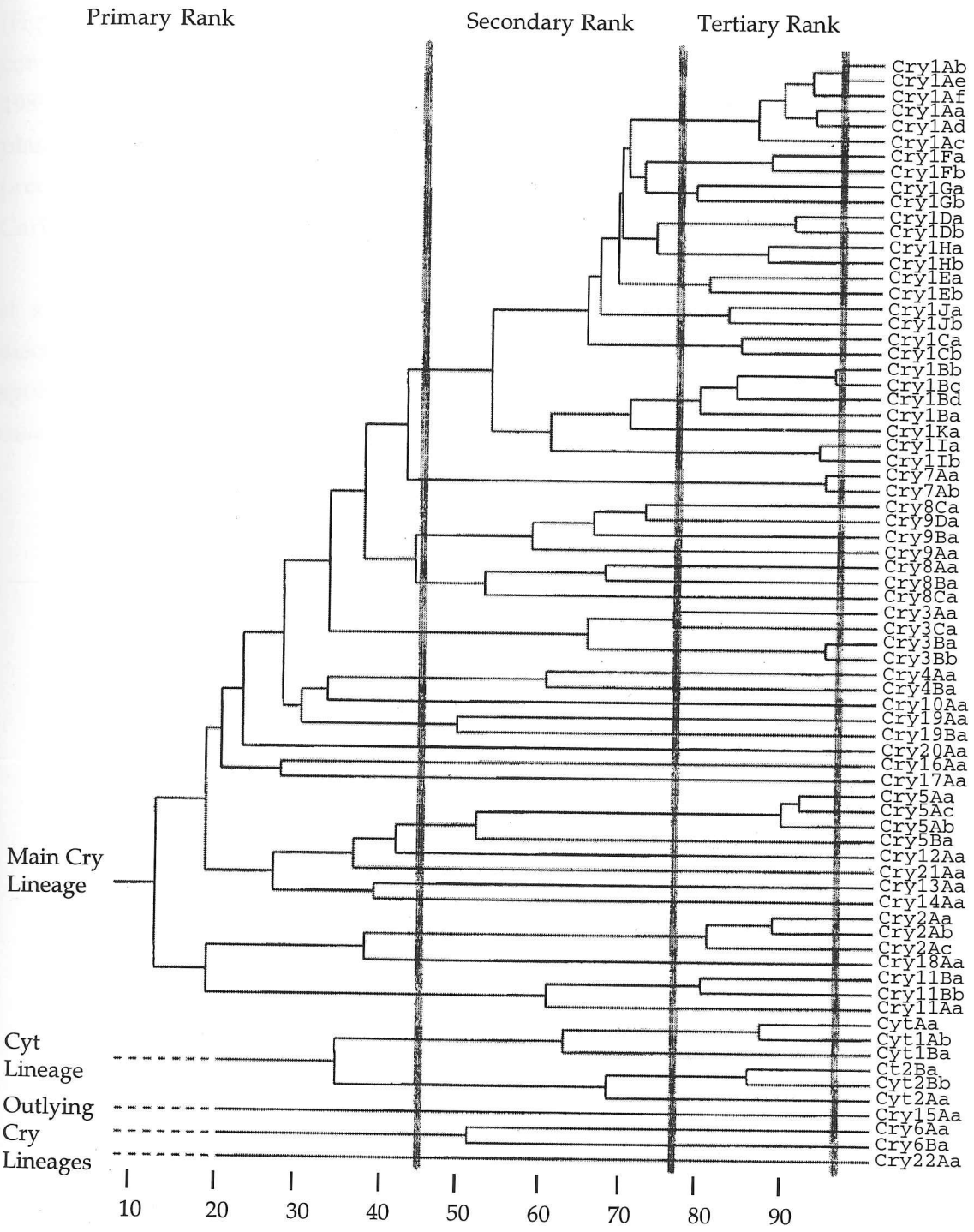


Figure 1.5 Phylogenetic tree demonstrating amino acid sequence identity among Cry and Cyt proteins.
Taken from Crickmore *et al.*, 1998. The grey vertical bars demarcate the four levels of nomenclature ranks. Based on the low percentage of identical residues and the absence of any conserved sequence blocks in multiple sequence alignments, the lower four lineages are not treated as part of the main toxin family, and their nodes have been replaced with dashed horizontal lines in the figure.

are confined to a few of the many different sized plasmids (1.4 MDa to about 150 MDa) (Figure 1.6) (Gonzalez *et al.*, 1981) often present in these isolates. Larger plasmids may contain two or more crystal genes in close proximity (Chambers *et al.*, 1991; Sanchis *et al.*, 1989) and in many cases the loss of crystal production can be traced to the loss of a single plasmid (Gonzalez *et al.*, 1981). Some Bt plasmids are also capable of conferring crystal production on recipient acrystalliferous Bt or *B. cereus* by conjugal transfer (Gonzalez & Carlton, 1982).

It has been reported that some δ-endotoxin genes reside on chromosomal DNA (Klier *et al.*, 1982; Kronstad *et al.*, 1983) although it may be possible that difficulties in discrimination between large (>150 MDa) plasmids and chromosomal DNA has led to spurious results. The recent identification of a Cry toxin in *Clostridium bifermentans* ssp. *malaysia* (Barloy *et al.*, 1996) has shown that the gene encoding it is chromosomally located.

Figure 1.6 Plasmid profiles of Bt subspecies.
The photograph is a composite of two photographs from Carlton and Gonzalez (1985). Courtesy of P.A. Koni (photo no. 235C3).

A *thuringiensis* HD2, serotype 1
B *thuringiensis* HD931, serotype 1
C *finitimus* HD3, serotype 2
D *alesti* HD4, serotype 3a
E *anduze*, serotype 3a
F *kurstaki* HD73, serotype 3ab
G *kurstaki* HD263, serotype 3ab
H *kurstaki* HD1, serotype 3ab
I *darmstadiensis* HD146, serotype 10
J *darmstadiensis* HD498, serotype 10
K *toumanoffi*, serotype 11a
L *toumanoffi* HD540, serotype 11a
M *kyushuensis* HD541, serotype 541
N *kyushuensis* HD571, serotype 541
O *thompsoni* HD542, serotype 12
P *pakistani* HD395, serotype 13
Q *pakistani* HD462, serotype 13
R *israelensis* 4QI, serotype 14
S *wuhanensis* HD525, no serotype

Bt δ-endotoxins are also often found in close association with mobile genetic elements such as insertion sequences and transposons (Lereclus *et al.*, 1984; Kronstad & Whiteley, 1984; Mahillon *et al.*, 1994; Léonard *et al.*, 1997) which are discussed in greater detail in Chapter 8. Such regions seem to allow gene transposition, duplication, rearrangement and modification (Lereclus *et al.*, 1983) and may well facilitate the transfer of δ-endotoxin genes not only between plasmid and chromosome DNA (Gonzalez *et al.*, 1981; Carlton & Gonzalez, 1985), but also between plasmids (Whiteley & Schnepf, 1986) serving a major adaptative function. The horizontal dissemination of *cry* genes as part of composite transposons on single conjugative plasmids within the *B. cereus-B. thuringiensis* family may therefore provide a mechanism whereby Bt can create new strains with novel activities and host spectra.

Regulation of transcription

Despite the fact that the δ–endotoxins are generally encoded on large, low copy number plasmids, extraordinarily high levels of crystal protein synthesis are obtained from these few gene copies. This and the precise co-ordination of crystal protein synthesis with sporulation are controlled by a variety of mechanisms occurring at the transcriptional, posttranscriptional and posttranslational levels that have provided an interesting system for the study of gene expression (Agaisse & Lereclus, 1995; Baum & Malvar, 1995).

Transcription of the δ-endotoxin genes is generally restricted to the stationary or sporulation phase of the growth cycle and only during δ-endotoxin synthesis (Klier *et al.*, 1982; Wong *et al.*, 1983; Ward & Ellar, 1986; Kostichka *et al.*, 1996). The only known exception is Bt ssp. *tenebrionis* where *cry3Aa* transcripts and crystal protein antigens could be detected in vegetatively growing cells (Sekar, 1988). The *cry1I* gene is also unusual in that it is expressed from early stationary phase (Kostichka *et al.*, 1996) and the protein product exported into the supernatant.

The *cry1Aa* gene is a typical example of a sporulation-dependent *cry* gene expressed only in the mother cell compartment of Bt. Two adjacent transcriptional start sites (BtI and BtII) have been mapped (Wong *et al.*, 1983) defining overlapping promoters that, used sequentially may be a way of ensuring continued high expression of δ-endotoxin genes. Transcription occurs from BtI during stages II and III of sporulation and from BtII during Stage III. Sporulation-specific RNA polymerases have been isolated for BtI (Brown & Whiteley, 1988) and for BtII (Brown & Whiteley, 1990) that have allowed the promoter sequence to be more accurately defined (Figure 1.7).

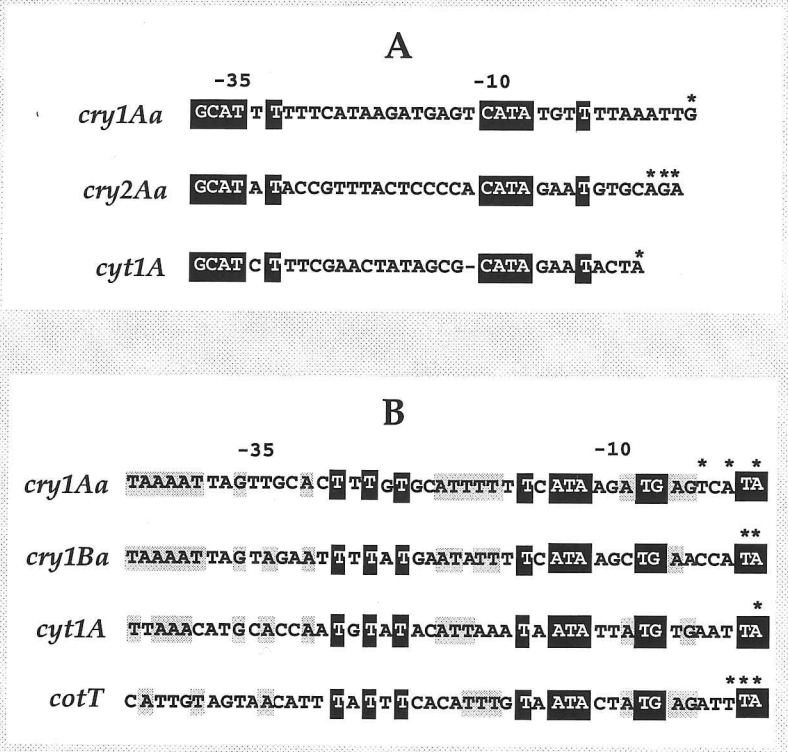


Figure 1.7 Alignments of the early/middle and middle/late sporulation-specific promoter sequences, recognised by σ^{35} (A) and σ^{28} (B) RNA polymerases, respectively.

The alignments are reproduced from Brown and Whiteley (1988 & 1990). Bases common to all of the σ^{35} or all of the σ^{28} promoters are outlined in black. Bases found in all but one of the σ^{28} sequences are shaded. Major transcription start sites for δ-endotoxin gene promoters *in vivo* and the observed *in vitro* start sites for the promoter of the spore coat protein gene *cot*T are indicated with asterisks.

Promoters BtI and BtII are recognised by sporulation-specific sigma factors σ^{35} and σ^{28} respectively (Brown & Whiteley, 1988 &1990) which show great similarity to those in *Bacillus subtilis* (Adams *et al.*, 1991; Stragier & Losick, 1990). Similar regions containing one or both types of promoters have been located upstream of several other *cry* genes (Ward & Ellar, 1986; Brizzard *et al.*, 1991; Yoshisue *et al.*, 1993a & 1993b). For example, the dipteran-active Bt ssp. *israelensis* produces five δ-endotoxins in its crystal; Cry4A, Cry4B, Cry10A, Cry11A and Cyt1A (Purcell, 1997). The complexity of the crystal composition made it difficult to clarify the mechanism regulating Bt ssp. *israelensis* crystal synthesis. Nevertheless, transcription of *cyt1A* was found to be initiated from promoters corresponding to BtI and BtII (Ward & Ellar, 1986; Waalwijk *et al.*, 1985). Yoshisue *et al.* (1993a & 1993b) showed that σ^{35} but not σ^{28} is responsible for transcription of *cry4A* and *cry4B* in Bt and that *cry4A* and *cry4B*-specific RNAs peak during mid-sporulation.

However, it was also reported that *cry4A* transcription is partially controlled by σ^H (a factor which functions primarily during stationary phase prior to septation) (Yoshisue *et al.*, 1995). Cry11A is expressed as the second gene of a three gene operon (Dervyn *et al.*, 1995), which encodes the previously identified 20-kDa protein (Adams *et al.*, 1989) and a newly identified 19-kDa protein (P19) (Dervyn *et al.*, 1995). It was demonstrated that transcription was regulated mainly by the σ^{35} sporulation-specific sigma factor, although σ^{28} also appears to play some role.

There are several examples that suggest that high mRNA levels further boost δ-endotoxin synthesis during sporulation. With a half-life of about 10 minutes, sporulation-specific mRNA is six times more stable than mRNA in vegetative cells (Glatron & Rapoport, 1972; Andrews *et al.*, 1982). There is some evidence that this stability arises from the presence of a potential stem-loop structure at the putative transcription terminator site (Wong & Chang, 1986), which is found in the transcripts from representatives of every δ-endotoxin pathotype (Whiteley & Schnepf, 1986; Ward & Ellar, 1986; Sekar *et al.*, 1987; Widner & Whiteley, 1989). The presence of this structure at the 3' end of the mRNA may protect it from exonuclease degradation and help to maintain high mRNA levels despite the low gene copy number.

In some species, mRNA levels are also increased by a DNA region upstream of the coding sequence. In the case of *cry3A*, a 1 kb DNA fragment located 400 base pairs upstream of the promoter enhances expression at the transcriptional level by increasing the number of *cry3A* specific transcripts during sporulation (Teixeira de Souza *et al.*, 1993). Interestingly, no potential coding sequences were found and so it was concluded that the region itself was important for enhancing transcription. The inclusion of the -35 region of the *cry3A* promoter within a potential cruciform structure formed by a perfect inverted repeat, suggests that the -35 region may be locked or unlocked by a regulator system. The 1 kb fragment also contains a putative cruciform structure and it was suggested (Teixeira de Souza *et al.*, 1993) that this DNA sequence is a binding site for a specific activator, which through local modification of the DNA structure, could unlock and expose the -35 region of the promoter thereby enhancing *cry3A* transcription.

Post translational effects on expression

Post translational effects have also been shown to play a part in increasing the production of some δ-endotoxins. Proteins possibly act as molecular chaperones either by facilitating toxin folding to a stable conformation or by protecting the unfolded toxin from proteolytic degradation (Visick & Whiteley, 1991).

Probably the best studied example is the effect of a 20-kDa protein from Bt ssp. *israelensis* on the production of the Cyt1A protein. In *E. coli*, expression of cloned *cyt1A* was observed to increase significantly when a region of DNA located 4 kb upstream of the *cyt1A*

promoter was present (McLean & Whiteley, 1987). This region was shown to carry a gene encoding a 20-kDa protein which appeared to enhance the production of Cyt1A (Adams *et al.*, 1989). It appeared that the 20-kDa polypeptide acted post-translationally to increase Cyt1A synthesis, occurred in small amounts in the Cyt1A inclusion and possibly protected the δ-endotoxin from proteolysis since it was found to be dispensable in mutants with reduced proteolytic ability (Visick & Whiteley, 1991). The related Cyt1B however, has no requirement for the 20-kDa protein for efficient expression, possibly reflecting its greater resistance to proteolysis (Koni & Ellar, 1993). The 20-kDa protein was also found to significantly improve net synthesis of Cry11A and promote its crystal formation (Wu & Federici, 1995). The improved production of proteins as diverse as Cry11A and Cyt1A indicated that the 20-kDa protein may be useful in facilitating the production of other proteins. This was found to be the case when experiments showed that the impact of extensive modifications to the protease-resistant domain of Cry1C, which were detrimental to the rate of production of protein and the formation of crystal, could be partially compensated for by co-expression with the 20-kDa protein (Rang *et al.*, 1996).

A similar observation has been noted for Cry2A. The *cry2A* gene appears as the third in a three gene operon (*orf1, orf2, cry2A*) (Widner & Whiteley, 1989). Expression of *cry2A* was dramatically reduced in the absence of *orf2* and no Cry2A inclusions were observed (Crickmore & Ellar, 1992). It has been postulated that Orf2 acts as a molecular chaperone, either assisting in Cry2A cuboidal crystal formation or by protecting the δ-endotoxin from proteolytic degradation (Crickmore & Ellar, 1992).

Effect of copy number

For engineered strains, the copy number of plasmids carrying the *cry* genes could determine the relative expression of the different toxin genes present. This is therefore an important consideration when constructing recombinant Bt strains harbouring several toxin genes.

The influence of copy-number on the expression of *cry* genes in Bt was examined by Arantes and Lereclus (1991). The *cry3A* gene from the coleopteran-specific Bt LM79 was cloned into each of three novel shuttle vectors constructed with low (4 per equivalent chromosome), intermediate (15) and high (70) copy number and the resulting plasmids introduced into a *cry* minus Bt strain by electroporation (Lereclus *et al.*, 1989). Bacteria harbouring the low copy number plasmid showed no significant difference in protein production (as estimated by the procedure of Lowry *et al.*, 1951) to those carrying vector alone. However, bacteria bearing intermediate and high copy number plasmids both produced about three times as much protein. The lack of toxin production with the low copy number plasmid suggested that the level of gene expression is not exclusively related to vector copy number, but is influenced by additional factors. The existence of a repressor

was hypothesised as one of these additional factors. It was proposed that the repressor is only detectable when the *cry* gene is present at a low copy number in the cell, is titrated, and then ineffective when the *cry3A* gene is cloned in a higher copy number plasmid.

There was no significant difference between the amounts of δ-endotoxin produced by the cells harbouring intermediate and high copy number plasmids, thus indicating that above fifteen copies of plasmid per chromosome, factors other than *cry* gene copy number limit the synthesis of crystal protein (Arantes & Lereclus, 1991). It may well be the case that the amount of sigma factor is limiting for toxin expression, as was observed in *B. subtilis*, where the introduction of a *spo*VG promoter on a multi-copy-number plasmid reduced expression of the chromosomal gene (Banner *et al.*, 1983).

1.6 Delta-Endotoxin Mechanism of Action

Larval susceptibility

Bacillus thuringiensis δ-endotoxins are gut poisons rather than contact insecticides and need to be ingested by the insect larvae in order to exert a toxic effect. The mechanism of action involves solubilisation of the crystal in the midgut, proteolytic processing of the protoxin by midgut proteases and binding of the activated toxin to midgut receptors followed by insertion into the apical membrane creating ion channels or pores. Larval susceptibility can be affected by any one, or a combination of these stages.

Different δ-endotoxins are active against different insect species but, even if an insect is susceptible, death may take anything from seven hours to seven days depending on the severity of toxin action in the larval gut (Heimpel & Angus, 1959; Dulmage, 1981; Miyasono *et al.*, 1994; Beegle & Yamamoto, 1992). The dose of toxin present and the presence or absence of spores may also have an effect (Dulmage, 1981; Miyasono *et al.*, 1994). However, as far as the use of Bt as an insecticide for crop protection is concerned, the important factor is not how long it takes for the insect to die, but how long it takes before it stops feeding and, for susceptible insects this is not long.

For the most sensitive insects ingestion of δ-endotoxin can lead to cessation of feeding caused by gut and general paralysis within ten minutes (Heimpel & Angus, 1959). The mixing of gut contents with the haemolymph causes a change in blood pH of 1-1.5 pH units which leads to death within seven hours. Less sensitive insects still undergo gut paralysis and stop feeding almost immediately after ingesting the toxin but there is no general paralysis or mixing of the gut contents with the haemolymph and death by starvation may take two to seven days (Heimpel & Angus, 1959). Other insects (e.g. *Anagasta kuehniella*) with a nearly neutral gut pH are susceptible to spore-crystal mixes only and death results from septicaemia after spore germination in the midgut.

δ–Endotoxin solubility

The insect gut consists of three regions, the fore, mid and hind gut (Figure 1.8). Since the foregut and hindgut are protected by an impermeable cuticle, the first area of exposed tissue encountered by the toxin is the midgut epithelium, the most likely site of δ-endotoxin action. It is in the midgut of many insects, including Lepidoptera (Berenbaum, 1980; Dow, 1984) and Diptera (Dadd, 1975; Lacey & Federici, 1979), that the crystals are met by some of the most alkaline conditions found in biological systems (up to pH 12 in some species (Berenbaum, 1980; Dow, 1984; Dadd, 1975; Lacey & Federici, 1979). These conditions are at least partially responsible for solubilisation of the δ-endotoxin inclusion and can be mimicked *in vitro* using high pH solutions such as ammonium carbonate, sodium carbonate and sodium hydroxide (Hannay, 1953; Angus, 1956; Lecadet, 1966, 1967; Bulla *et al.*, 1977, 1981; Huber *et al.*, 1981). Efficient solubilisation is highly important to δ-endotoxin potency. For some δ-endotoxins, solubilisation *in vitro* prior to ingestion by the insect can increase toxicity (Jaquet *et al.*, 1987; Andrews *et al.*, 1982; Bulla *et al.*, 1981) implying inefficient *in vivo* solubilisation. Since an insect will only be susceptible if the δ-endotoxin inclusion is solubilised in its gut, δ-endotoxin solubility can be regarded as the initial determinant of specificity.

For complete solubilisation of some crystal types at pH values between 9 and 10, the addition of reducing agents is required. This enhanced solubility implied that disulphide bonds are important for the maintenance of crystal formation (Dastidar & Nickerson, 1979). This hypothesis was confirmed when it was shown that cysteine residues of Cry1-class protoxins are involved in interchain, symmetrical disulphide bonds that stabilise the crystal lattice and account for the unusual solubility of these proteins (Bielot *et al.*, 1990). The distribution of cysteine residues is biased towards the highly conserved C-terminal half of the toxin and the first few N-terminal residues (Brousseau & Masson, 1988; Höfte & Whiteley, 1989). It is therefore thought that these residues may play an important structural role in crystal formation and stability (discussed in section 1.7).

The Cry2, Cry3, Cry10A and Cry11A protoxins lack this C-terminal region and their solubility reflects this fact. For example Cry11A and Cry2 inclusions are soluble at pH>10.5 without the addition of reducing agents (Nishiitsuji-Uwo *et al.*, 1980; Thomas & Ellar, 1983; Couche *et al.*, 1987; Nicholls *et al.*, 1989; Federici *et al.*, 1990). The method of crystal solubilisation is unknown in these cases, but it may be significant that the genes which encode Cry2 and Cry11A proteins are associated with accessory proteins (Orf1, P19, Orf2) (Widner & Whiteley, 1989; Dervyn *et al.*, 1995; Wu *et al.*, 1991). The composition of the δ-endotoxin crystal can also affect its solubility. Aronson *et al.*, (1995) demonstrated that the absence of Cry1Ab protoxin markedly reduced the solubility of all the inclusion proteins, presumably because of the increased disulphide cross-linking among protoxins.

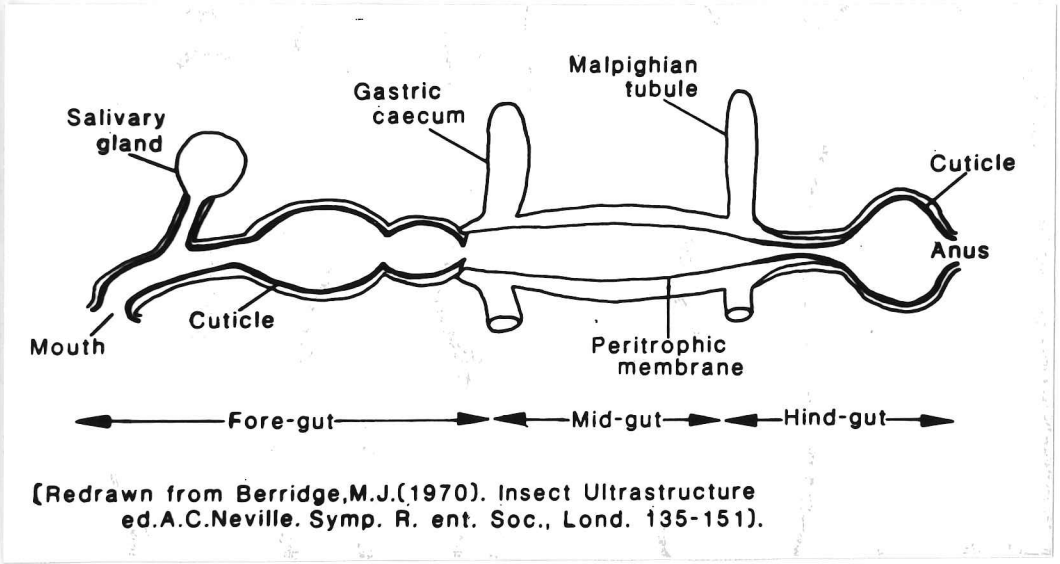


Figure 1.8a Diagrammatic representation of the insect gut (photo no. 247C6).

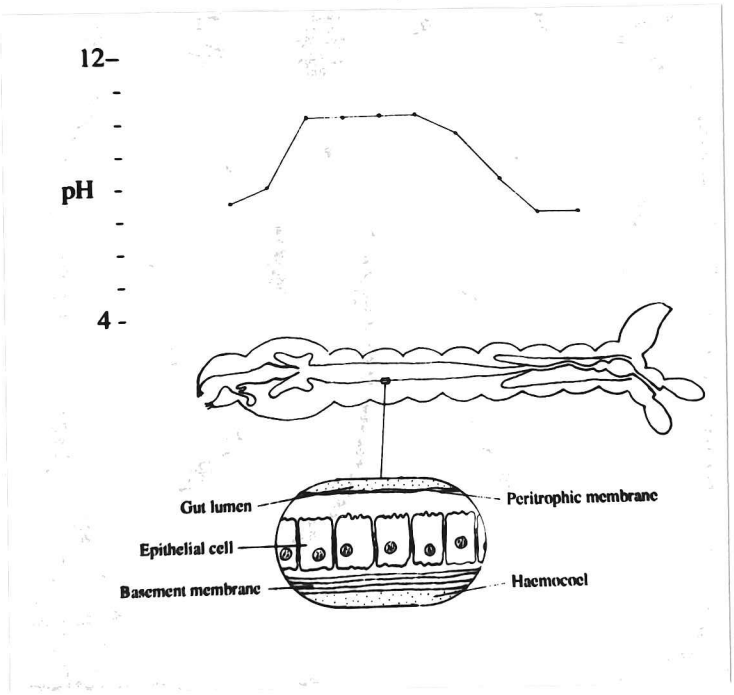


Figure 1.8b Diagrammatic representation of *Culex pipiens* mosquito larva showing the morphology and pH profile along the gut length.
Courtesy of T. J. Sawyer (photo no. 856A5).

Proteolytic activation of δ-endotoxins

Solubilised Bt δ-endotoxins exist as biologically inactive protoxins until they are converted to functional toxins by proteolytic activation. The lack of toxicity displayed by solubilised protoxins compared to solubilised and protease treated protoxins was demonstrated when both forms were injected into the haemocoel of lepidopteran larvae (Lecadet & Martouret, 1967), or exposed to cultured insect cell lines (Johnson, 1981). In the protease-rich environment of the insect gut, the processes of inclusion solubilisation and proteolytic activation almost certainly occur simultaneously, the activation step being carried out by naturally occurring insect proteases. The midgut lumens of lepidopteran and dipteran larvae contain a variety of proteases including serine-proteases which are optimally active in the alkaline pH of the insect midgut. Dipteran larvae have been shown to be rich in chymotrypsin-like and trypsin-like enzymes (Yang & Davies, 1971; Kunz, 1978) which have been demonstrated to activate protoxins molecules *in vitro*. Midguts of susceptible coleopteran larvae contain mainly cysteine proteases which are active under the mildly acidic conditions of the coleopteran gut environment (Thie & Houseman, 1990; Purcell *et al.*, 1992). Toxin activation can be reproduced *in vitro* by solubilising the protoxin crystals in buffers and then incubating with either prepared larval gut extracts or commercially available serine-proteases such as trypsin or chymotrypsin (Thomas & Ellar, 1983; Haider *et al.*, 1986). However, prolonged *in vitro* activation leads to the production of smaller polypeptide products with lower or no toxicity (Lüthy *et al.*, 1982).

Proteolytic processing of the δ-endotoxins can occur at either the C-terminal end (Chestukhina *et al.*, 1982; Aronson *et al.*, 1986), the N-terminal end (Nicholls *et al.*, 1989; Carroll *et al.*, 1997) or both (Nagamatsu *et al.*, 1984; Höfte *et al.*, 1986), depending on the δ-endotoxin class. The 120 to 140-kDa lepidopteran-specific δ-endotoxins are activated by removal of 500-600 residues at the C-terminus and 27-29 residues at the N-terminus, to leave an active moiety of 60 to 70-kDa derived from the amino terminal half of the protoxin. Deletion analysis of several *cry1* type protoxin genes confirms that the insecticidal activity resides in the N-terminal half of the protein and that any further truncation of the active polypeptide can abolish its *in vivo* toxicity (Schnepf & Whiteley, 1985; Höfte *et al.*, 1986). In contrast, the 72-kDa dual-specific Cry2A proteins undergo little or no proteolysis at their C-termini, but undergo more extensive N-terminal processing (Cry2A is activated by the removal of 144 residues from the N-terminus (Nicholls *et al.*, 1989)). The diptericidal Cry4A and Cry4B δ-endotoxins appear to be activated in a similar way to Cry1 toxins; the activity being located in the N-terminal half (Angsuthanasombat *et al.*, 1991; Chilcott & Ellar, 1988; Chungjatupornchai *et al.*, 1988). Studies suggest that Cry11A is cleaved into two halves of 30 and 35-kDa (Chilcott & Ellar, 1988), but it is not known if both are active fragments.

The Cyt1A and Cyt1B δ-endotoxins both require proteolytic activation, mainly in the N-terminus (Koni & Ellar, 1994), to express their toxicity (Drobiewski & Ellar, 1989; Knowles *et al.*, 1992; Koni & Ellar, 1994; Chilcott & Ellar, 1988).

δ–Endotoxin receptors

Bt δ-endotoxins appear ideally adapted to damage a susceptible insect using the insect's own physiology. The insect's highly alkaline gut pH solubilises the crystal to release inactive protoxin (possibly in non-susceptible vertebrates and invertebrates with lower gut pHs the crystal would pass through the gut intact and therefore be available again for ingestion), which is then converted to an active stomach poison by the insect's own gut proteases. Once activated it is thought that toxin binding proteins (receptors) located in the brush border membranes of insect midgut cells (Wolfersberger, 1990) direct the δ-endotoxins to bind to the membrane and exert their toxic effect. Studies using brush border membrane vesicles (BBMV) prepared from insect larval midguts (Hofmann *et al.*, 1988a & 1988b; Van Rie *et al.*, 1989 & 1990a) and *in vitro* studies using insect cell lines (Knowles & Ellar, 1986; Haider & Ellar, 1987a) have provided convincing evidence for the existence of such receptors which may well play a role in the normal physiological processes of the epithelial cells (Regev *et al.*, 1996).

The insect epithelium is essentially a monolayer of cells, resting on a basement membrane of connective tissue and surrounded by a discontinuous layer of muscle. These midgut cells are covered by a membrane layer of chitin, protein and glycosylated protein, called peritrophic membrane (Brandt *et al.*, 1978; Adang & Spence, 1982; Berner *et al.*, 1983; Peters, 1992), which is thought to offer some protection from abrasive food particles and microbial infection. Several reports have suggested that peptides with a maximum molecular weight of 66-kDa could pass easily across the peritrophic membrane, and this would enable activated toxin, but not protoxin, to gain access to the midgut epithelium (Yunnnovitz *et al.*, 1986; Wolfersberger *et al.*, 1986).

Receptor binding is the key factor in specificity. In most cases, a positive correlation between toxicity and binding to the specific receptor proteins has been observed, although several exceptions are known (Garczynski *et al.*, 1991; Hofmann *et al.*, 1988a, 1988b; Lee *et al.*, 1992; Van Rie *et al.*, 1989, 1990a; Wolfersberger, 1990). Two different types of insect protein have been identified as receptors for Cry toxins, the 120-kDa aminopeptidase N Cry1Ac toxin-binding protein purified from the brush border vesicles of *Manduca sexta*, *Heliothis virescens* and *Lymantria dispar* (Gill *et al.*, 1995; Knight *et al.*, 1994; Sangadala *et al.*, 1994; Valaitis *et al.*, 1995) and the 210-kDa cadherin-like glycoprotein Cry1Ab toxin-binding protein purified from *M. sexta* membrane (Vadlamundi *et al.*, 1995). An aminopeptidase N from *Bombyx mori* has also been proposed as a candidate for the receptor of the Cry1Aa toxin (Yaoi *et al.*, 1997). Aminopeptidase N is thought to be

anchored in the membrane by a lipid, probably glycosylphosphatidylinositol (GPI) anchor (Takesue *et al.*, 1992). Different δ-endotoxins may bind to different receptors and it is possible that these may also include other cell-surface components such as glycolipids.

The Cyt toxins appear to differ from the Cry toxins in that they display broad spectrum cytolytic activity *in vitro*, lysing most eukaryotic cells tested (Thomas & Ellar, 1983). This has been attributed to their high hydrophobicity and ability to bind to unsaturated phospholipids (Thomas & Ellar, 1983; Gill *et al.*, 1987; Drobniewski & Ellar, 1989; Knowles *et al.*, 1992). However, the greatest demonstrable *in vitro* activity of these toxins is towards mosquito cells (Chilcott & Ellar, 1988) and they are specific to dipteran larvae *in vivo*. There is considerable evidence that unsaturated phospholipids are the receptor for Cyt toxins (Gill *et al.*, 1987; Drobniewski & Ellar, 1988, 1989). If this is the case, then the higher abundance of these lipids in dipteran membranes (Fast, 1966; Lukonen *et al.*, 1973; Jenkin *et al.*, 1976) could account for the preferential dipteran toxicity of the Cyt toxins *in vitro* and *in vivo* (Li *et al.*, 1996).

The development of insect resistance to Bt δ-endotoxins has been proposed to be due either to altered protoxin activation by midgut proteases, loss of specific midgut receptors or their binding capacity (Ferré *et al.*, 1995; Van Rie *et al.*, 1990b). Reversal of such resistance was reported to occur when exposure to Bt was stopped for many generations (Tabashnik *et al.*, 1993). Regained sensitivity was linked to restored δ-endotoxin-binding capacity and higher biotic fitness, indicating that the toxin receptors might also play a role in normal physiological processes of the epithelial cells. Although the molecular basis of this phenomenon is still unclear, it may indicate a delicate equilibrium between larval biotic fitness and the antagonising selectable pressure exerted by exposure to specific δ-endotoxins. Cross resistance can also occur when selection with one δ-endotoxin or set of δ-endotoxins reduces susceptibility to others (Tabashnik *et al.*, 1996). Cross resistance is most likely when toxins share key structural features (Tabashnik *et al.*, 1996) and provides further evidence that certain receptors may bind a range of δ-endotoxins.

δ-Endotoxin mechanism of action

The δ-endotoxin mechanism of action has been studied *in vitro* by using a variety of systems including insect cell lines (Murphy *et al.*, 1976), whole insect guts (Harvey & Wolfersberger, 1979) and BBMVs purified from insect guts (Hofmann *et al.*, 1988a). Various mechanisms have been proposed (Griego *et al.*, 1979; Sacchi *et al.*, 1986; Lüthy *et al.*, 1982; Faust *et al.*, 1974; Wolfersberger, 1991) but there is mounting evidence that Bt toxins kill cells in a biphasic mechanism of colloid osmotic lysis proposed by Knowles and Ellar in 1987.

In this model, specific binding involves two steps, one that is reversible and one that is irreversible. After the toxin binds the receptor, it is thought that there is a change in the

toxin conformation allowing toxin insertion into the membrane (Knowles, 1994). Recent data suggests that toxicity correlates with irreversible binding (Liang *et al.*, 1995) which might be related to insertion of the toxin into the membrane but could also reflect tighter interaction of the toxin with the receptor. Oligomerisation of the toxin follows insertion and this oligomer then forms a non-specific pore or lesion in the membrane (Li *et al.*, 1991; Lorence *et al.*, 1995; Sacchi *et al.*, 1986) (Figure 1.9). Trans-epithelial ion gradients are collapsed by leak through the pore and there is an osmotically driven influx of water, resulting in cell swelling and eventual lysis (Knowles & Ellar, 1987; Ellar *et al.*, 1990).

Both Cry and Cyt toxins have been shown to form pores in planar lipid bilayers and liposomes under certain conditions (Slatin *et al.*, 1990; English *et al.*, 1991; Schwartz *et al.*, 1993; Smedley *et al.*, 1997; Knowles *et al.*, 1989, 1992). Experiments done *in vitro* with cultured insect cell lines showed that both Cyt and Cry toxins elicited a graded efflux of molecules from cells and that cytotoxicity could be delayed or inhibited by osmotic protectants such as raffinose that are too large to penetrate the toxin-induced pores (Knowles & Ellar, 1987; Haider & Ellar, 1987a; Drobniewski *et al.*, 1987). Using this approach, a pore diameter was calculated for several Bt δ-endotoxins based on time courses in which osmotic protectants were used to inhibit or delay cytolysis as assessed by Trypan blue uptake. The use of light scattering to analyse changes in the permeability of BBMV treated with toxin (Carroll & Ellar, 1993) has provided further evidence for the colloid osmotic lysis model, by showing that Cry1Ac-treated BBMV from *M. sexta* displays increased permeability for cations, anion and neutral solutes as well as water. Cry1B, which is not toxic towards *M. sexta* had no effect.

As previously discussed in Section 1.3, midgut cell lysis also seems to be the underlying mechanism of toxicity associated with the newly discovered class of Bt vegetative insecticidal proteins, Vips (Yu *et al.*, 1997). This same mechanism has also been connected with cholesterol oxidase, another insecticidal protein with acute toxicity (Purcell *et al.*, 1993). A common mechanism of midgut cell lysis therefore seems to be the strategy adopted by the most effective insecticidal proteins. This is perhaps not surprising when it is considered that, since the midgut is the site of crucial functions like digestion and absorption, interference in these functions by disrupting gut cells will generate a very potent toxic effect.

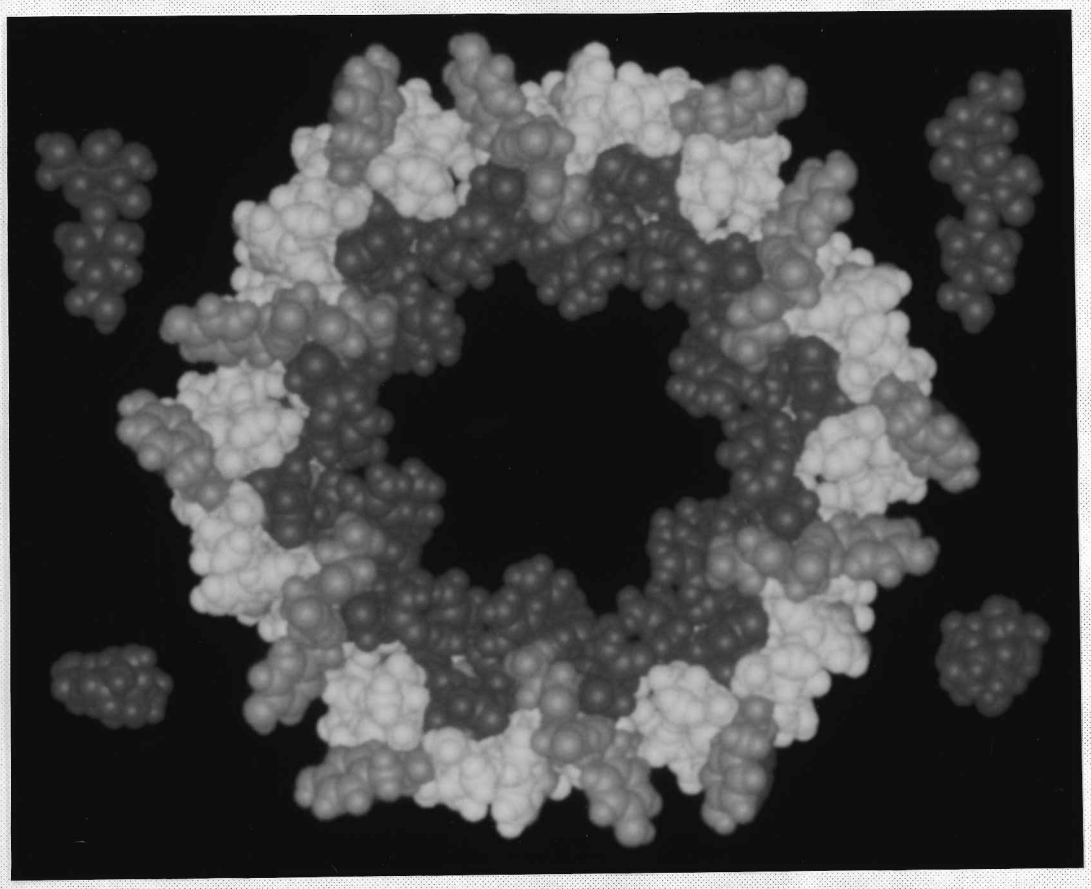


Figure 1.9 Suggested model of the Bt toxin pore (reproduced from Hodgeman & Ellar, 1990). Space filling model showing proposed pore consisting of a hexamer of putative helices 4 and 5 of Cry1Ab; to the left are two views of raffinose, (a sugar which has been shown to block the pore, Knowles & Ellar, 1987), and to the right, two views of sucrose, (a sugar which can diffuse through the pore). Hydrophilic residues are shown in dark blue, (including histidine, asparagine and glutamine), uncharged hydrophilic residues in light blue, (including serine, theronine, glycine and tyrosine), and hydrophobic residues in white.

1.7 Three Dimensional Structures

The crystal structures of Cry3A (coleopteran-specific) (Figure 1.10) and Cry1Aa (lepidopteran-specific) toxins have been reported (Li *et al.*, 1991; Grochulski *et al.*, 1995) and has contributed greatly to a better understanding of the structure-function relationships of Cry δ-endotoxins. The Cry3A protoxin has a molecular mass of 70-kDa and does not contain the large carboxyl-terminal extension contained in the Cry1Aa toxin. The crystal structure of the Cry1Aa toxin was determined from the activated toxin fragment. Both toxins share 36% amino acid identity and the two structures show high overall similarity (Grochulski *et al.*, 1995). Both are globular molecules containing three distinct domains connected by single linkers.

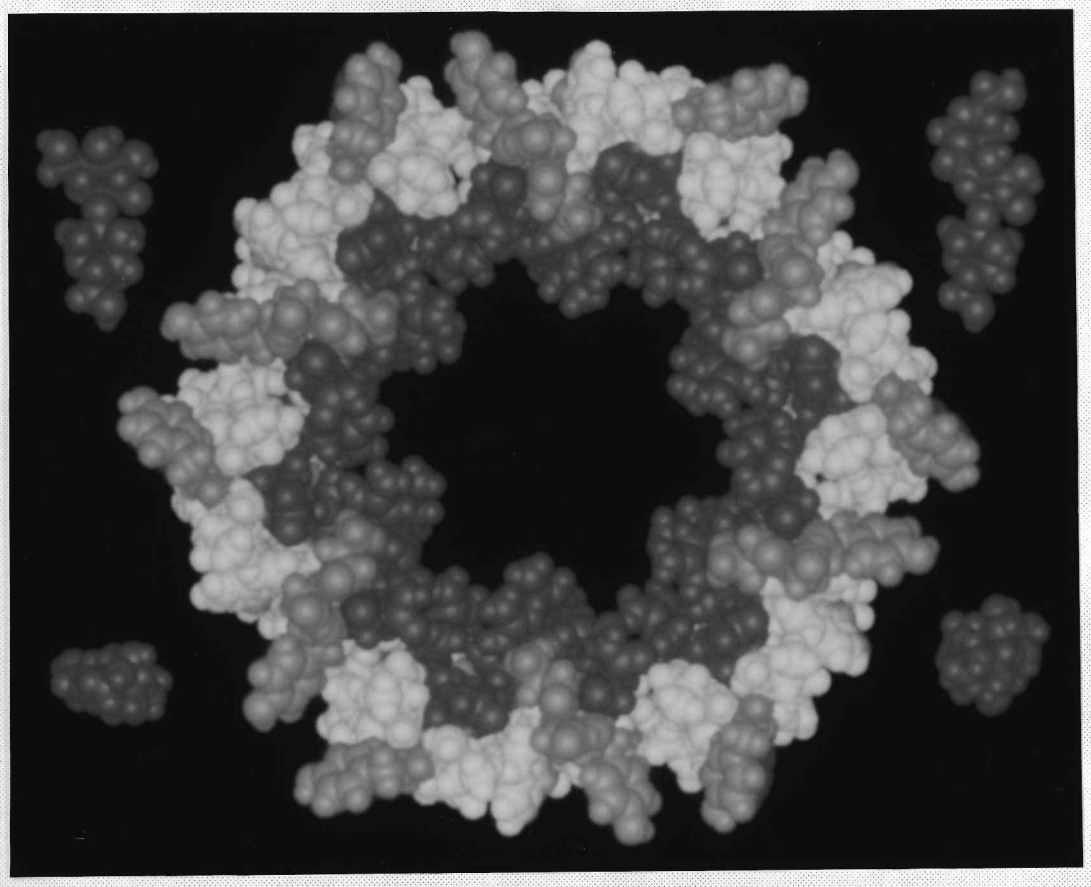


Figure 1.9 Suggested model of the Bt toxin pore (reproduced from Hodgeman & Ellar, 1990). Space filling model showing proposed pore consisting of a hexamer of putative helices 4 and 5 of Cry1Ab; to the left are two views of raffinose, (a sugar which has been shown to block the pore, Knowles & Ellar, 1987), and to the right, two views of sucrose, (a sugar which can diffuse through the pore). Hydrophilic residues are shown in dark blue, (including histidine, asparagine and glutamine), uncharged hydrophilic residues in light blue, (including serine, theronine, glycine and tyrosine), and hydrophobic residues in white.

1.7 Three Dimensional Structures

The crystal structures of Cry3A (coleopteran-specific) (Figure 1.10) and Cry1Aa (lepidopteran-specific) toxins have been reported (Li *et al.*, 1991; Grochulski *et al.*, 1995) and has contributed greatly to a better understanding of the structure-function relationships of Cry δ-endotoxins. The Cry3A protoxin has a molecular mass of 70-kDa and does not contain the large carboxyl-terminal extension contained in the Cry1Aa toxin. The crystal structure of the Cry1Aa toxin was determined from the activated toxin fragment. Both toxins share 36% amino acid identity and the two structures show high overall similarity (Grochulski *et al.*, 1995). Both are globular molecules containing three distinct domains connected by single linkers.

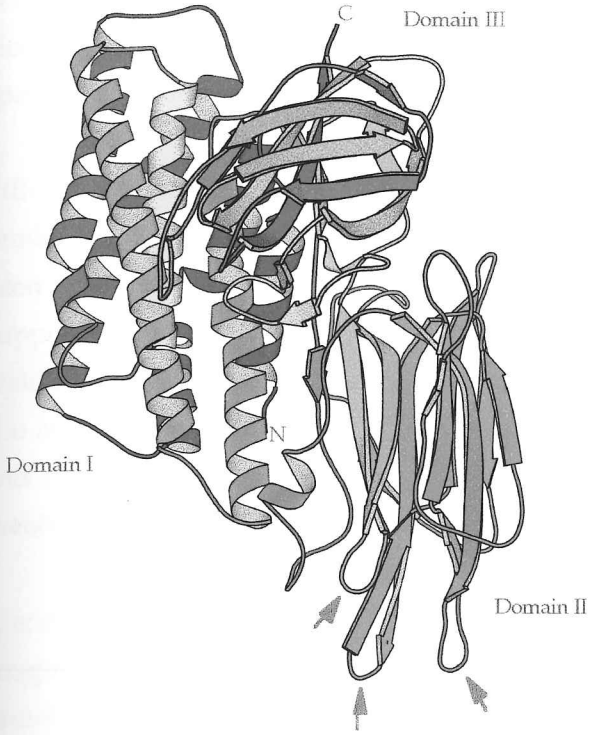


Figure 1.10 The three dimensional structure of Cry3A (Li *et al.*, 1991; refined and updated by Dr. J. Li).
The polypeptide pathway is indicated by colouring the chain in the rainbow order, from red at the N-terminus to blue at the C-terminus. The three domains are: I, a seven helix bundle (upper left): II, a three-sheet assembly (bottom) and: III, a β-sandwich (upper right). The three exposed loops in Domain II are indicated by arrows. Photograph courtesy of Dr. D. J. Ellar.

Domain I at the N-terminus of the molecule is a seven α–helix bundle in which the central helix (α5) is surrounded by six outer helices. (Figure 1.10) Five of these helices (α3–α7) are long enough to span a 30Å thick membrane bilayer suggesting that domain I may be involved in channel formation in the membrane. Point mutations in the region encoding the central α5 helix of the Cry1Aa toxin drastically affect toxicity without affecting binding to larval midgut vesicles (Wu & Aronson, 1992) and support this theory. However, the conformation of domain I would have to be reversed during pore-formation to provide a hydrophobic outer layer for contact with the lipid bilayer and a hydrophobic inner surface to form an aqueous channel. This model requires a flexible Cry toxin structure capable of unfolding and penetrating into target insect membranes. This flexibility may be obtained in Cry3A by inter-helical proteolytic nicking of domain I (Li *et al.*, 1991). Confirmation that nicking takes place between helix α3 and α4 in Cry3A has recently been obtained (Carroll *et al.*, 1997). Proteolysis has also been found to occur between putative helices α3 and α4 in Cry1B when treated with gut extract from a susceptible insect species (*P. brassicae*) (J. Carroll, personal communication). Proteolysis within putative domain I regions of both Cry4B (Angsuthanasombat *et al.*, 1993) and Cry2A (Nicholls *et al.*, 1989) has also been reported with retention of toxic activity and no fragment separation.

Domain II consists of three antiparallel β–sheets with similar topologies packed around a hydrophobic core (Li *et al.*, 1991) (Figure 1.10) and represents the most divergent

part of the structure between Cry3A and Cry1Aa (Grochulski *et al.*, 1995). This domain has been described as a specificity-determining domain since reciprocal hybrid genes between closely related toxins (Cry1Aa and Cry1Ac) resulted in chimeric toxins with altered specificity (Ge *et al.*, 1991; Schnepf *et al.*, 1990).

Three exposed loops in domain II (see Figure 1.10) show the largest structural differences between Cry3A and Cry1Aa (Grochulski *et al.*, 1995) and are conspicuously reminiscent of the antigen recognition site of an immunoglobulin. These loops have therefore been proposed as the possible receptor recognition region. This theory has since been supported by mutational analysis. Mutations in loop 1 of Cry1Aa demonstrated that these residues are essential for binding to the brush border membrane of *Bombyx mori* midgut cells (Lu *et al.*, 1994). Mutations to both loops 1 and 2 of Cry1C were shown to affect toxicity and specificity (Smith & Ellar, 1994) and loops 1 and 3 of Cry3A have been found to be involved in irreversible binding to *Tenebrio molitor* midgut membranes (Wu & Dean, 1996).

Previous work (Knowles & Ellar, 1986; Knowles *et al.*, 1991) has indicated that the specificity of Cry1Ac toxin is determined by the nature of the carbohydrate attached to the receptor. Comparison of the Cry toxin structure with the crystal structure of VMO-1 (a protein composed entirely of β-prism structure found in the vitelline membrane of hens eggs) provided the first indirect evidence that this β-prism structure is a carbohydrate-targeting domain as VMO-1 is thought to function as an oligosaccharide binding protein (Shimuzu *et al.*, 1994). More recently the solving of the X-ray structure of the plant lectin, jacalin (Sankaranarayan *et al.*, 1996) has provided more compelling evidence to support the role of Domain II in receptor recognition. Each subunit of this tetrameric protein contains the β-prism fold which can be superimposed on the corresponding Cry toxin domain. Jacalin binds specifically to a tumour associated T-cell disaccharide and the X-ray structure of the form with bound methyl-α-D galactose shows that the carbohydrate binding site of jacalin is composed of the exposed loops that connect the folded β-strands.

At the C-terminus, domain III is a sandwich of antiparallel β-sheets (Figure 1.10). The function of domain III is still the subject of debate. It has been proposed that it stabilises the toxin by providing protection from proteolysis (Li *et al.*, 1991). However, recent reports suggest that it may be involved in channel function as a voltage sensor, since conservative mutations in this region reduced toxicity without reducing binding (Chen *et al.*, 1993). Mutations of two acidic residues in β–23 of Cry4A (Nishimoto *et al.*, 1994) that form hydrogen bonds with arginine residues of β-17 caused an unstable conformation as judged by digestion with proteases. This supported the idea that both β-sheets are important for the proper folding of the toxin (Grochulski *et al.*, 1995).

Several lines of evidence indicate that domain III may also be involved in receptor binding. The construction of chimeric proteins has shown that domain III of Cry1C is a significant determinant of specificity to *Spodoptera exigua* and *Mamestra brassicae* (Bosh *et*

al., 1994). Also it has been demonstrated that domain III exchanges between Cry1Ac and Cry1Aa toxins affect binding to different *L. dispar* midgut receptors (Lee *et al.*, 1995) and mutations in this region in Cry1Ac resulted in proteins that bound poorly to the *M. sexta* and *H. virescens* toxin-binding proteins leading to extensive loss of toxicity for both insects. More recently a triple mutation in domain III of Cry1Ac completely eliminated binding to the carbohydrate GalNAc and to aminopeptidase N in *M. sexta* (Burton, 1998).

Höfte and Whiteley (1989) identified five highly conserved regions among the sequences of Cry toxins. The locations of these regions in the three-dimensional structures of Cry3A and Cry1Aa are the same. The high degree of conservation of these sequence blocks and their important structural location (at central positions of each domain or involved in inter-domain contacts) led Li *et al.*, (1991) to propose that Cry toxins which possess these blocks would share a similar structure, one of globular toxins composed of three structural domains. More recently, a complete amino acid sequence alignment of all Cry proteins so far published showed most of them to have the same five conserved blocks plus, three additional blocks lying outside the active core in the carboxyl-terminal halves of sequences with more than 1,000 residues (Schnepf *et al.*, 1998) (Figure 1.11).

Experimental data from several laboratories supports the hypothesis that δ-endotoxins have a modular structure and have shown that domains from Cry proteins are structurally independent. Domain I (Von-Tersch *et al.*, 1994; Walters *et al.*, 1993) and helix α5 peptides (Cummings *et al.*, 1994; Gazit *et al.*, 1994), expressed independently, retain their ability to form cation channels in planar lipid bilayers and the exchange of sequence segments within domains II and III resulted in specificity changes (Ge *et al.*, 1991; Schnepf *et al.*, 1990).

An explanation of the independence of the three domains was given by Bravo (1997) after her extensive phylogenetic analysis of the Cry toxin family showed that the biocidal activity of the Cry toxins has resulted from two fundamental evolutionary processes. Firstly, the independent evolution of the three functional domains and secondly, domain swapping among different toxins. She proposed that it is by these two processes that such a large family of proteins has been generated with similar modes of action but very different specificities.

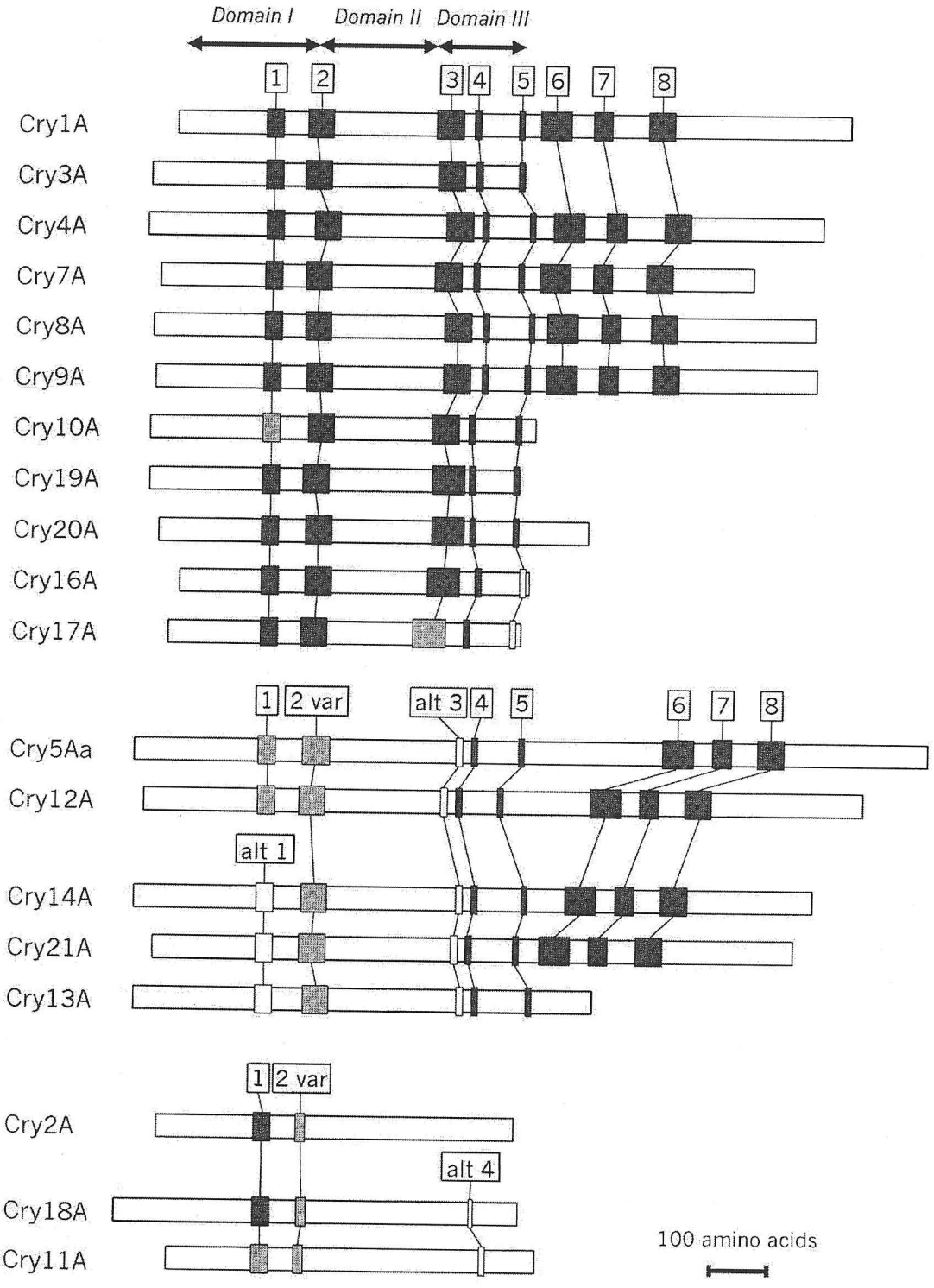


Figure 1.11 Positions of conserved sequence blocks among Cry proteins.
Taken from Schnepf *et al.*, 1998. Sequence blocks are shown as dark grey, light grey, or white to indicate high, moderate, or low degrees of homology, respectively, to the consensus sequence for each conserved block. Highly conserved sequences conform to the consensus sequence at 75% or more of its positions. Variant sequences (Var) conform to the consensus sequence of the highly conserved group at 50 to 75% of the positions and Alternate blocks (Alt) are derived from groups of proteins having a consensus sequence over that sequence block that differs from the corresponding highly conserved sequence at more than half its positions. The lengths of each protein and the conserved blocks within them are drawn to scale.

1.8 Aims and Objectives of this Investigation

The identification of novel diptericidal proteins may be valuable both in the development of pesticides with novel specificities and also in delaying or even preventing resistance development to currently used formulations. It has been found that combining various toxins with similar specificities in one recombinant cell may prevent or delay the onset of resistance (Georghiou *et al.*, 1993, 1997). Therefore this work started with an investigation into the polypeptide profiles of two diptericidal strains of Bt ssp. *fukuokaensis*: 84-I (Yu *et al.*, 1991) and 17A (Karamanlidou *et al.*, 1991) in the hope that novel δ-endotoxins would be identified.

Should novel Cry proteins be identified it is hoped that one or more of the novel genes may be cloned. Subsequent characterisation of their protein products may enable differences in size and amino acid sequence to be correlated with potential differences in stability, solubility and activity towards various insects. Information such as this will strengthen our knowledge of Cry toxins and their mode of action.

Investigations into the phylogeny of any new *cry* gene sequences are also planned. It has recently been suggested that phylogenetic analysis of Bt δ-endotoxins could be used to identify potential activities and direct strategies to create novel chimeric toxins with different specificities (Bravo, 1997). Use of such analysis may therefore indicate the direction of any further work.

Finally, a variety of Bt toxin genes have been found in association with transposable elements (Kronstad & Whiteley, 1984; Lereclus *et al.*, 1984) and toxin genes are often found in regions of DNA containing other toxin genes and/or accessory proteins. It is therefore proposed that the genetic context of any *cry* genes cloned during the course of this project should be investigated.

Two

Chapter 2

Materials and Methods.

2.1 Materials

2.1.1 Antibiotic Stocks

Ampicillin (100 mg/ml) dissolved in dH_2O.

Chloramphenicol (10 mg/ml) dissolved in 96% ethanol.

Antibiotic stocks were prepared at the concentrations stated above by filtering through 0.22 µm filters before storage at 20°C.

2.1.2 Bacterial Strains

The *E. coli* strains used in this work were JM109 (Promega) {*end*A1, *thi*, *gyr*A96, *hsd*R17 (r_K^- m_K^+), *rel*A1, *sup*E44, (*lac-pro*AB), [F', *tra*D36, *pro*AB$^+$, *lacI*^{q}ZΔM15]} and TG2 (*supE*, *hsd*Δ*5*, *thi* Δ(*lac-proAB*) Δ(*srl-recA*) 306::Tn10 (*tet*^{r}) F'[*traD36*, *proAB*$^+$, *lacI*^{q}, *lac*ZΔM15]). *E. coli* strains were normally grown at 37°C for 4-16 hours in LB medium (Sambrook *et al.*, 1989).

Bacillus thuringiensis strains used were subspecies *fukuokaensis* 84-I-I-13 (Yu *et al.*, 1991), *fukuokaensis* 17A (Ormilia) (Karamanlidou *et al.*, 1991) and subspecies *israelensis* IPS78/11 (a plasmid-cured mutant of Bt subspecies *israelensis* lacking δ-endotoxin genes (Crickmore & Ellar, 1992). Bt strains were normally grown at 30°C for 4-16 hours in LB medium.

2.1.3 Bacterial Growth Media

CCY Sporulation Medium	Stewart *et al.*, 1981
Luria-Bertani Medium	Sambrook *et al.*, 1989
PWYE	Ellar and Posgate, 1974

2.1.4 Buffers and Solutions

All buffers and solutions are referred to in subsequent Chapters by the titles given. All were made with MilliQ purified, deionised water (dH_2O), stored at -20°C, 4°C or room temperature and were diluted with dH_2O before use where appropriate. Where DTT or PMSF was included, these were added immediately prior to use.

5x Calf intestinal alkaline phosphatase buffer: 50 mM Tris pH 8.3, 5 mM $MgCl_2$, 5 mM $ZnCl_2$

Coomassie stain: 0.1% Coomassie blue, 10% methanol, 10% acetic acid

Denaturation Buffer: 0.5 M NaOH, 1.5 M NaCl

DIG Buffer 1: 0.1M Maleic acid, 0.15 M NaCl, pH 7.5

DIG Buffer 2: 0.3% Tween 20 in DIG buffer 1

DIG Buffer 3: 0.1 M Tris, pH 9.5

10x DNA loading buffer: 5x TBE buffer, 50% glycerol, 0.25% bromophenol blue

10x Restriction Endonuclease buffer (supplied by Pharmacia): 10x One-Phor-All

Lowry solution: 2% Na_2CO_3, 0.1 M NaOH, 0.02% sodium tartrate, 0.01% cupric sulphate

Neutralisation buffer: 0.5 M Tris pH 7.5, 1.5 M NaCl

PBS buffer (supplied in tablet form from Oxoid): 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 140 mM NaCl, 2.7 mM KCl, pH 7.4

Prehybridisation solution: 5x SSC buffer, 2% skim milk powder, 1% SDS, 0.1% N-laurylsarcosine

2x Protein loading buffer: 50 mM Tris pH 7.5, 10% glycerol, 1% SDS, 1 mM EDTA pH 8, 0.05% bromophenol blue, 50 mM DTT, 5 mM PMSF

Semi-dry blot buffer: 39 mM glycine, 48 mM Tris, 0.0375% SDS (w/v), 20% methanol (v/v)

20x SSC: 3 M NaCl, 0.3 mM tri-sodium citrate

50x TAE buffer: 2 M Tris 5.7% acetic acid, 50 mM EDTA pH 8

10x TBE buffer: 0.89 M Tris, 0.89 M boric acid, 20 mM EDTA pH 8

10x TBS buffer: 0.1 M Tris pH 7.4, 0.15 M NaCl

10x TE buffer: 0.1 M Tris pH 8, 10 mM EDTA pH 8

2.1.5 Chemicals

All chemicals used were Analytical Reagents or Ultra Pure where possible.

Acetic acid	Fisons
Acetone	Fisons
ATP	BDH Chemicals
Agarose MP	Boehringer Mannheim
Ammonium persulphate	Fisons
Bacteriological agar	Oxoid
Boric acid	Rhone-Pôulenc
5-bromo-4-chloro-3-indolyl-β-D-galactopyranose (X-Gal)	Sigma
Bromophenol blue	Sigma
Caesium chloride	Melford Laboratories
Chloroform	Fisons
4-chloro-1-naphthol	Sigma
Coomassie Brilliant Blue R	National Diagnostics

Cupric sulphate pentahydrate	Fisons
Deoxynucleoside triphosphates (dNTPs)	Pharmacia
Dithiothreitol (DTT)	Melford Laboratories
Ethanol	BDH Chemicals
Ethidium bromide	Sigma
Ethylenediaminetetra-acetic acid (EDTA)	BDH Chemicals
EGTA	Sigma
Folin and Ciocalteu's reagent	Sigma
Glycerol	Fisons
Glycine	BDH Chemicals
Haemoglobin (bovine)	Sigma
Hydrochloric acid	Fisons
Hydrogen peroxide	Fisons
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Sigma
N-laurylsarcosine	Sigma
Magnesium chloride	Fisons
Magnesium sulphate	Fisons
Maleic acid	BDH Chemicals
β-Mercaptoethanol	BDH Chemicals
Methanol	Fisons
Mineral oil	Sigma
Parafin (liquid)	Sigma
Phenol	BDH Chemicals
Phenyl methane sulphonyl fluoride (PMSF)	Sigma
Phosphate buffered saline (PBS) tablets	Oxoid
Potassium chloride	Fisons
Propan-2-ol	Fisons
Sodium carbonate	Fisons
Sodium chloride	Fisons
Sodium dodecyl sulphate (SDS)	BDH Chemicals
Sodium hydroxide	May and Baker Ltd.
Sucrose	BDH Chemicals
Skim milk powder	Beta Lab
N, N, N', N' Tetramethyl ethylene diamine (TEMED)	Sigma
Thioglycolic acid	BDH
Trichloroacetic acid	Sigma
Tris hydroxymethyl aminoethane (Tris)	Sigma
Triton X-100	Fisons

Tryptone	Difco
Tween-20	Sigma
Yeast Extract	Beta Lab
Zinc chloride	BDH Chemicals

2.1.6 Computer Programes

The computer programes used to manipulate DNA sequence were from the Sequence Analysis Software Package of the Genetics Computer Group (GCG) (University of Wisconsin, Biotechnology Centre).

2.1.7 Enzymes

Calf Intestinal Alkaline Phosphatase	Boehringer Mannheim
Klenow fragment	Boehringer Mannheim
Lysozyme (egg white)	Sigma
Peroxidase-linked anti-rabbit IgG antiserum	Sigma
Proteases (trypsin, proteinase K)	Sigma
Restriction Endonucleases	Pharmacia & Boehringer Mannheim
Ribonuclease A	Sigma
T7 DNA polymerase	Pharmacia
T4 DNA ligase	Boehringer Mannheim

2.1.8 Miscellaneous Materials

Aedes aegypti eggs	Zeneca Agrochemicals
30% Acrylamide/0.3% bis acrylamide	Severn Biotech
Digoxygenin 3' tailing kit	Boehringer Mannheim
Bovine Serum Albumin Fraction V (BSA)	Sigma
Geneclean II kit	Bio 101 Inc.
Gene Elute spin column	Supel Co.
Nested deletion kit	Pharmacia
Nitrocellulose filter roll	Schleicher & Schuell
Nylon filter roll (positively charged)	Boehringer Mannheim
Perfectprep™ plasmid preparation kit	5 Prime —> 3 Prime
Photographic developer and fixer	Kodak
Polaroid 665 and 667 film	Polaroid
Problot membrane	Applied Biosystems
SDS-7 protein size markers	Sigma
Tip 20 DNA preparation kit	Qiagen
XR X-ray film	Fuji

2.1.9 Oligonucleotides

Oligonucleotides were produced by the Oligonucleotide Synthesis Facility, Department of Biochemistry, University of Cambridge. Freeze-dried oligonucleotides were resuspended in sterilised 0.1x TE for use in PCR and hybridisations, and sterile distilled water for use in sequencing. The stock solutions were then aliquoted and stored at -20°C.

2.1.10 Plasmids

pGEM-3Zf(+) (Stratagene) was used as a general cloning vector. The Bt-*E. coli* shuttle vector used was pSVP27A (Crickmore & Ellar, 1992). Plasmids were stored at either 4°C in the short term or -20°C in the long term in TE pH8.

2.2 Molecular Biology Methods

2.2.1 Small Scale Isolation of Plasmid DNA from *E. coli* cells

The methods used to isolate plasmid DNA from *E. coli* strains JM109 and TG2 were as follows.

2.2.1a Standard minipreparations of plasmid DNA were carried out using the Alkaline Lysis method (Birnboim and Doly, 1979). RNaseA was included in Solution I at a concentration of 250 µg/ml.

2.2.1b Preparation of plasmid DNA for sequencing was carried out using the Perfect prep™ plasmid DNA preparation kit from 5 Prime —> 3 Prime, Inc. according to the manufacturers protocol.

2.2.1c Preparation of plasmid DNA for nested deletions was carried out using either the Perfect prep™ plasmid DNA preparation kit from 5 Prime —> 3 Prime, Inc. or Tip 20s from Qiagen, both according to the manufacturers protocol.

2.2.2 Isolation of Plasmid DNA from Bt cells

2.2.2a Small scale isolation of plasmid DNA from Bt cells was carried out by the alkaline lysis method stated in 2.2.1a with the following modification. Solution I contained 10 mg/ml of lysozyme and after the pellet had been resuspended in this solution the tube was incubated at 37°C for 30 minutes before the addition of solution II.

2.2.2b Plasmid DNA prepared for use in cloning procedures was purified using caesium chloride/ethidium bromide continuous gradients subsequent to alkaline lysis (Birnboim and Doly, 1979; Sambrook *et al.*, 1989). Native plasmids were extracted and purified from *Bacillus thuringiensis* using a method based on the alkaline lysis method incorporating the following modifications:

1. One litre of pre-warmed LB medium in a 2 litre flask was inoculated with 2 ml of an overnight culture and grown with shaking overnight (12-16 hours).

2. Cells were harvested by centrifugation in a 500 ml wide-mouthed bottle at 5000x *g* for 10 minutes and the cell pellet resuspended in Buffer I to make a total volume of 27 ml.
3. 3 ml of Buffer I containing 10 mg/ml of lysozyme was added and the suspension incubated at 37°C for 30 minutes.
4. The suspension was then placed on ice and 60 ml of Buffer II was added. The suspension was mixed by gentle inversion and returned to the ice for another 5 minutes.
5. 40 ml of 3 M potassium/5 M acetate was added, the suspension mixed again by inversion and placed on ice for a further 5 minutes.
6. The mixture was then centrifuged at 10,000x *g* for 5 minutes at room temperature.
7. The supernatant was filtered through muslin into a fresh 250 ml wide mouth bottle and 0.6 volumes of propan-2-ol was added, mixed gently and left for 10 minutes at room temperature to precipitate the nucleic acids.
8. The mixture was then centrifuged at 10,000x *g* for 30 minutes at room temperature and the supernatant was discarded. The pellet was dissolved in 7 ml of 10x TE buffer in a 15 ml centrifuge tube before 8 g of caesium chloride and 0.6 ml of ethidium bromide (10 mg/ml) was added to the mixture.
9. This mixture was centrifuged at 4000x *g* for 5 minutes and the protein-ethidium bromide precipitate removed from the top of the tube.
10. The solution from this centrifugation was transferred to a 13.5 ml ultracentrifuge tube, filled with liquid paraffin and balanced against a second tube. The tubes were then centrifuged at 45000x *g* for 47-70 hours.
11. After centrifugation the lower plasmid band was separated from any higher linear DNA band using a silicanised Pasteur pipette (Sambrook *et al.*, 1989).
12. The ethidium bromide was removed from the plasmid solution by several extractions using NaCl saturated propan-2-ol.
13. The plasmid solution was then diluted five-fold using 1x TE before the DNA was precipitated with 2.5 volumes of 96% ethanol in a 30 ml Corex tube and spun at 10,000x *g* for 15 minutes.
14. The resulting pellet was washed twice in ethanol (80%), vacuum dried and resuspended in up to 0.5 ml 1x TE.

2.2.3 Restriction Endonuclease Digestion

Restriction endonuclease digestions were carried out in the "One-Phor-All" Pharmacia buffers supplied unless otherwise stated. Routinely, 0.5-1 µl of restriction endonuclease was added to a DNA buffer mix containing 0.2-1.0 µg of DNA to give a total volume of 20 ml. Incubation was for 1-2 hours at 37°C.

2.2.4 Agarose Gel Electrophoresis

The agarose concentration used depended on the size of the DNA fragments to be separated. Routinely 0.8% gels (w/v) were used. Gels were made up with either 1x TBE or 1x TAE buffer and ethidium bromide added to a final concentration of 0.5 µg/ml. Samples were loaded with 1/6 volume of gel loading buffer. Bacteriophage Lambda DNA restricted with *Hin*dIII was run on gels as molecular weight markers.

Electrophoresis was carried out at 1-10 volts per centimeter of gel length for 1-4 hours using the same running buffer as in the gel. The bands were then visualised on a UV transilluminator at 254 nm.

2.2.5 DNA Fragment Isolation

The DNA sample was electrophoresed in the presence of ethidium bromide on an agarose gel of an appropriate concentration (Sambrook *et al.*, 1989) using a TAE buffer system. The DNA was then briefly visualised under ultra-violet light (365 nm) and the appropriate band excised with a scalpel. The DNA was purified from the gel slice by Geneclean II or using a Supel Co. Gene Elute spin column as described by the manufacturers.

2.2.6 5' Dephosphorylation of Linearised Vectors

The DNA sample was made up to 0.75-0.8 volumes of the final reaction volume with dH_2O before 0.2 volumes of the final reaction volume of 5x calf intestinal alkaline phosphatase buffer was added and mixed. The reaction was then made up to the final reaction volume with 1 unit/µl calf intestinal alkaline phosphatase and mixed by pipetting before being incubated at 37°C for one hour. The reaction was stopped on addition of one tenth volume of 200 mM EGTA and incubating at 65°C for ten minutes.

The efficiency of dephosphorylation of linearised vector was tested by using a small fraction of the preparation to compare the number of colonies seen with that from an equivalent amount of non-dephosporylated linearised vector after both had been subjected to a ligation reaction and then used to transform *E. coli* JM109.

2.2.7 DNA 3' End-filling

A DNA sample with recessed 3' ends was made up to 0.4-0.5 volumes of the final reaction volume with dH_2O before 0.1 volumes of each of the four dNTPs and 10x Klenow DNA polymerase buffer was added to a final concentration of 1x and mixed. The dNTP stocks were at a concentration of 2.5 mM. The volume was made up to the final reaction volume with 5 units/ml Klenow DNA polymerase and mixed before being left at room temperature for thirty minutes.

2.2.8 DNA Ligation

The DNA sample was made up to 0.7 volumes of the final reaction volume with dH_2O before 0.1 volumes each of One-Phor-All buffer and fresh 10 mM ATP was added and mixed. The reaction was then made up to the final reaction volume with T4 DNA ligase and mixed by pipetting before being incubated at 15°C for 16-20 hours or at room temperature for 2-5 hours.

2.2.9 3' End Labeling of Oligonucleotide Probes

A one hundred picomole sample of oligonucleotide probe was made up to 0.8-0.9 volumes of the final reaction volume with dH_2O before adding the following solutions from the DIG-labeling kit: 4 µl tailing buffer, 5 µl $CoCl_2$ solution, 1 µl DIG-dUTP solution, 1 µl dATP solution and 1 µl (50 units) terminal transferase. This mixture was incubated at 37°C for 30 minutes then placed on ice. DIG-tailed oligonucleotide was precipitated with 2.5 µl LiCl (4 M) and 75 µl pre-chilled absolute ethanol. The solution was mixed well and left at -70°C for 30 minutes before centrifugation at 10,000x *g* for 10 minutes. The pellet was washed with 50 µl cold ethanol (70% v/v), dried *in vacuo* and dissolved in an appropriate volume of sterile dH_2O and stored at -20°C. Incorporation of the nucleotide analogue was checked by carrying out trial dot-blots together with a labelled standard of known concentration according to the manufacturers instructions.

2.2.10 Transfer of DNA to Nylon Membranes

DNA was transferred from the agarose gel onto positively charged nylon membrane (Boerhinger Mannheim) for 16-20 hours as described by Southern (1975) following the manufacturers instructions (Boehringer Mannheim; "The DIG System User's Guide for Filter Hybridisation"). The agarose gel was soaked in denaturation solution (1.5 M NaCl, 0.5 M NaOH) twice for 15 min with gentle rocking, then twice for 15 min with rocking in neutralising solution (0.5 M Tris-HCl, pH 7.5, 3M NaCl). A platform was assembled in a glass dish filled with 2x SSC (20x: 3M NaCl, 0.3M Na_3 citrate). A wick of Watman 3MM filter paper was saturated in 2x SSC, laid over the platform with the gel placed on top. The gel was then overlaid with a sheet of nylon membrane (cut to the same size as the gel), 3 sheets of 3MM paper, a stack of paper tissues and a 100 g weight and left to transfer overnight (Sambrook *et al.*, 1989). The filter was then briefly washed in 2x SSC to remove pieces of agarose and the DNA fixed to the filter by UV cross-linking for 90 seconds at 365 nm each side.

2.2.11 Hybridisation of Oligonucleotides to Immobilised DNA

DNA hybridisations using digoxygenin-labeled probes (DIG) were performed using the methods described by the manufacturer (Boehringer Mannheim; "The DIG System User's

Guide for Filter Hybridisation"). Following Southern blotting the nylon membrane was sealed in a plastic bag with 20 ml of prehybridisation solution containing 5x SSC, 2% (w/v) skimmed milk powder; 0.1% (w/v) N-lauryl sarcosine; 0.02% (w/v) SDS and incubated at hybridisation temperature for at least one hour. Hybridisation was carried out with 5-10 ml of the same buffer at the same temperature in a sealed bag for 16-20 hours with 1-10 pmol/ml of the DIG-labeled oligonucleotide probe. The filter was then washed twice for 15 minutes at room temperature with 2x SSC; 0.1% SDS (w/v) and twice for 5 minutes at the hybridisation temperature with 1x SSC; 0.1% SDS (w/v). The filter was then either air-dried or used directly for chemiluminescent detection of hybridised DNA (Section 2.2.12).

2.2.12 Chemiluminescent Detection

The buffers used for the chemiluminescent procedure were described in Section 2.2.11. The filter was washed briefly (1-5 minutes) in Washing buffer. It was then pre-blocked in Buffer 2 with shaking at room temperature for 30 minutes and then probed for a further 30 minutes in 50 ml of the same buffer with 5 µl of anti-DIG-AP conjugate. The filter was then washed 2x 15 minutes with 100 ml Washing buffer and equilibrated 2-5 minutes in 20 ml of Buffer 3 before being incubated with 1 ml of Lumigen CSPD substrate at 37°C for 30 minutes. The filter was exposed to Fuji XR film for 30-60 minutes and developed as described in Section 2.2.13.

2.2.13 Autoradiography

Autoradiography was performed using Fuji XR film and Genetic Research Instrumentation cassettes. Autoradiographs were developed using Kodak developer and fixer.

2.2.14 Polymerase Chain Reaction

PCR analysis of Bt strains was carried out in the following manner using protocols developed by Dr. S. Gash. Frozen suspensions of *Bacillus thuringiensis* spores were grown up on LB agar plates overnight at 30°C. Cells were scraped off each plate into 500 µl aliquots of dH_2O, resuspended by brief vortexing and frozen on dry ice. The cell suspension was then boiled for 10 minutes, briefly vortexed and again snap-frozen on dry ice. The cells were then boiled for a second time and the suspension centrifuged at 13000x *g* for 1 minute. The supernatant was collected transferred to a fresh tube and subsequently stored at -20°C for future use.

In general the following conditions were used for PCR analysis of Bt strains. The final volume for each reaction was 100 µl with the final concentration of reagents being:-

Guide for Filter Hybridisation"). Following Southern blotting the nylon membrane was sealed in a plastic bag with 20 ml of prehybridisation solution containing 5x SSC, 2% (w/v) skimmed milk powder; 0.1% (w/v) N-lauryl sarcosine; 0.02% (w/v) SDS and incubated at hybridisation temperature for at least one hour. Hybridisation was carried out with 5-10 ml of the same buffer at the same temperature in a sealed bag for 16-20 hours with 1-10 pmol/ml of the DIG-labeled oligonucleotide probe. The filter was then washed twice for 15 minutes at room temperature with 2x SSC; 0.1% SDS (w/v) and twice for 5 minutes at the hybridisation temperature with 1x SSC; 0.1% SDS (w/v). The filter was then either air-dried or used directly for chemiluminescent detection of hybridised DNA (Section 2.2.12).

2.2.12 Chemiluminescent Detection

The buffers used for the chemiluminescent procedure were described in Section 2.2.11. The filter was washed briefly (1-5 minutes) in Washing buffer. It was then pre-blocked in Buffer 2 with shaking at room temperature for 30 minutes and then probed for a further 30 minutes in 50 ml of the same buffer with 5 µl of anti-DIG-AP conjugate. The filter was then washed 2x 15 minutes with 100 ml Washing buffer and equilibrated 2-5 minutes in 20 ml of Buffer 3 before being incubated with 1 ml of Lumigen CSPD substrate at 37°C for 30 minutes. The filter was exposed to Fuji XR film for 30-60 minutes and developed as described in Section 2.2.13.

2.2.13 Autoradiography

Autoradiography was performed using Fuji XR film and Genetic Research Instrumentation cassettes. Autoradiographs were developed using Kodak developer and fixer.

2.2.14 Polymerase Chain Reaction

PCR analysis of Bt strains was carried out in the following manner using protocols developed by Dr. S. Gash. Frozen suspensions of *Bacillus thuringiensis* spores were grown up on LB agar plates overnight at 30°C. Cells were scraped off each plate into 500 µl aliquots of dH_2O, resuspended by brief vortexing and frozen on dry ice. The cell suspension was then boiled for 10 minutes, briefly vortexed and again snap-frozen on dry ice. The cells were then boiled for a second time and the suspension centrifuged at 13000x *g* for 1 minute. The supernatant was collected transferred to a fresh tube and subsequently stored at -20°C for future use.

In general the following conditions were used for PCR analysis of Bt strains. The final volume for each reaction was 100 µl with the final concentration of reagents being:-

$MgCl_2$	2 mM
Promega Taq Polymerase reaction buffer	1x
Each Primer	0.25 mM
Template	15% (v/v)
Taq Polymerase	2.5 Units
Deoxynucleoside triphosphates	0.3 mM

The reactions were overlaid with mineral oil prior to PCR. A Hot-start PCR methodology was used routinely where the $MgCl_2$ was added separately to the PCR reaction mixture once both the $MgCl_2$ and the reaction mixture had been heated separately to 85°C. Cycling conditions are given at the appropriate section in the text.

Where PCR was used to generate DNA for use in cloning procedures, the template plasmid DNA was isolated using the Perfect prep™ plasmid DNA preparation kit (2.2.1b). It was this template DNA that was used to Hot-start the PCR.

2.2.15 DNA Sequencing

DNA sequencing was carried out by Mr. John Lester (Sequencing facility, Dept. of Biochemistry, University of Cambridge) using a model 373 Automated DNA sequencer (Applied Biosystems, Inc.) together with reagents and protocols supplied by Applied Biosystems, Inc. The primers used are stated in the appropriate section of the text. The DNA was prepared using the Perfect prep™ plasmid DNA preparation kit (2.2.1b).

2.2.16 Transformation

Cells were transformed by electroporation (Dower *et al.*, 1988) using a Biorad Gene Pulsar and Pulse controller.

2.2.16a Electroporation of *Escherichia coli* cells: Cells were grown to an OD_{600} of between 0.5-1.0, and chilled on ice for 30 minutes before harvesting. The cells were washed twice in ice cold sterile distilled water then again in ice cold sterile 10% glycerol before being resuspended in ice-cold 10% glycerol to a final culture concentration of five-hundred-fold. 50 µl aliquots were stored at -80°C until use.

Electroporation was performed using 50 µl of cell suspension in a 0.2 cm gap curvette at 25 µF, 1.5 kV, 1000 Ω for strain JM109. Electroporation efficiency was found to be $1x10^9$ transformants /µg.

Within one minute of electroporation 1 ml LB broth was added, the cells transferred to a sterile universal tube and incubated at 37°C with shaking for one hour before plating onto LB agar plates containing ampicillin (100 µg/ml).

2.2.16b Electroporation of *Bacillus thuringiensis* cells: Bt strain IPS78/11 was also transformed by electroporation using a method based on Bone and Ellar (1989).

A 1 cm^2 patch of bacteria was grown up on an LB agar plate overnight at 30°C. 0.5 ml of sterile distilled water was added to the plate and the bacteria spread evenly using a glass spreader. The plate was then incubated at 30°C for a further 2-4 hours until a thin lawn of cells was evident. 2 ml of cold sterile water was added to the plate, the cells resuspended using a fine glass spreader and transferred to a 1.5 ml Eppendorf tube. The cells were pelleted by centrifugation for 1 minute at 15,000x g, the supernatant discarded and the pellet resuspended in 1 ml of cold sterile 10% (w/v) sucrose. The cells were centrifuged again for 1-2 minutes at 15,000x *g* and the pellet resuspended in 120 µl of cold sucrose per transformation.

120 µl of the resuspended cells were transferred into 1 ml capacity electroporation curvettes with a gap of 4 mm and the DNA added (100-500 µg in 2-5 µl 1x TE). The curvette was kept on ice for 5 minutes before and after the DNA was added. The cells were transformed by electroporation using the following parameters; 2 kW, 400 Ω and 25 µF.

Within one minute of electroporation 1 ml LB broth was added, the cells transferred to a sterile universal tube and incubated at 30°C with shaking for one hour before plating onto LB agar plates containing chloramphenicol (10 µg/ml).

2.3 Protein Biochemistry

2.3.1 *Bacillus thuringiensis* Inclusion Preparation

Bt strains were streaked from frozen glycerol stocks onto CCY plates and grown at 30°C for 36-60 hours until spores and inclusions were observed by phase contrast microscopy. 1-3 colonies which were seen to have produced visibly normal spores and inclusions were then heat-activated in 1 ml of sterile water at 70°C for 30 minutes to kill any vegetative cells and promote subsequent synchronous growth. After cooling, the suspension was used to inoculate 250 ml of a rich medium (PWYE) and incubated overnight at 30°C with shaking (200 rpm). Sporulation was then induced by transfer of 1ml of cells to 1000 ml of pre-warmed minimal salts medium (CCY). All cultures were grown at 30°C in an orbital incubator (200 rpm) until approximately 95% of cells had released spores (approximately 36-48 hours).

NaCl was added to a final concentration of 1 M NaCl to inhibit proteolysis and the culture stored at 4°C for at least 1 hour. The spore/crystal mix was then harvested as described by Thomas and Ellar (1983), the final pellet being resuspended in 1/50 of the initial culture volume (to prevent aggregation of the spores and crystals). The spore/crystal mix was then separated on discontinuous sucrose gradients. 10 ml of 2.55 M, 8 ml of 2.4 M, 8 ml of 2.30 M and 5 ml of 1.97 M sucrose in Gradient Buffer were used for the native strains and 11 ml of 2.30 M, 10 ml of 2.10 M and 10 ml of 1.97 M for the cloned proteins. Centrifugation was carried out using a Beckman L5-50B ultracentrifuge for 16 hours at

37,000x *g* at 4°C. Crystals were then harvested, washed three times in ice-cold distilled water and stored in aliquots in distilled water at -20°C.

2.3.2 Rapid *Bacillus thuringiensis* Inclusion Isolation ("Quick Prep")

A Bt culture was streaked onto a CCY agar plate and incubated at 30°C for 30-60 hours until sporulation had occurred (see Section 2.8). Several colonies were transferred into 1 ml of 0.5 M NaCl, the tubes sonicated in a sonic water bath and then vortexed thoroughly. The mix was centrifuged at 13,000x *g* for five minutes and the supernatant discarded. The pellet was resuspended in 100 µl of 1% SDS/0.01% β-mercaptoethanol, boiled for 5 to 10 minutes in a water bath, then centrifuged for ten minutes at 13,000x *g*. The supernatant was removed and stored at -20°C until analysis by SDS-PAGE (Section 2.3.4).

2.3.3 Protein Estimation

Protein concentrations were determined by a method based on Lowry *et al.* (1951), using bovine serum albumin fraction V as a standard. Protein samples were made up to 0.6 ml in a 15 ml test tube before 3 ml of freshly prepared Lowry solution was added, mixed and left for 10 minutes. 0.3 ml of 50% Folin and Ciocalteu's phenol reagent was added, mixed and left for a further 30 minutes. Optical density at 750 nm was then determined against a control containing no protein. Assays were done in triplicate with a BSA range of 10-100 µg.

2.3.4 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis

SDS-PAGE was conducted as described by Laemmli and Favre (1973), using a discontinuous buffer system as modified by Thomas and Ellar (1983). Protein samples were solubilised by incubation at 100°C for 10 minutes in 2x SDS-PAGE loading buffer before being applied to a 13% (w/v) polyacrylamide gel with a 5% (w/v) polyacrylamide stacking gel. Electrophoresis was carried out at 20 mA until the dye front had entered the separating gel, and subsequently at 30 mA. Protein size markers were Sigma SDS-7 with added β-galactosidase. Proteins were visualised by incubating the gel in Coomassie stain with gentle shaking for 2 hours and destained in 10% methanol/10% acetic acid. Appropriate gels were then photographed (Section 2.4).

2.3.5 Immunoblot Analysis

Proteins separated by SDS-PAGE were electrophoretically transferred to nitrocellulose based paper by the method of Towbin *et al.* (1979) using an LKB Transblot semi-dry blotter. A blotting stack was assembled on the anode in the following order: 3 sheets of Whatman 3 MM paper, nitrocellulose paper, acrylamide gel and finally 3 more

sheets of Whatman 3 MM paper. All components of the stack were soaked in semi-dry blot buffer before use.

After transfer, non-specific binding was blocked with blocking buffer (3% BSA [w/v] dissolved in TBS) overnight with agitation. The blot was then incubated for one hour with δ-endotoxin rabbit antiserum (in fresh blocking buffer (1:500 dilution) raised against the relevant protein or proteins.

Excess antibody was removed with 5 washes of TBS (100 ml each) before incubation with peroxidase-conjugated goat anti-rabbit IgG antiserum (1:2000 dilution). Incubation was continued for one hour with agitation followed by removal of excess antibody as before. Bound antibodies were detected using 4-chloro-l-naphthol (30 mg) dissolved in methanol (10 ml) diluted in TBS (50 ml) with H_2O_2 (20 µl) (Hawkes *et al.*, 1982). Once the blots were developed the reaction was stopped by several changes of tap water. The blot was then stored in distilled water in darkness before being photographed.

2.3.6 δ-Endotoxin Inclusion Solubilisation

Purified δ-endotoxin inclusions were solubilised at 1-2 mg/ml by incubation at either 37°C or 4°C for 1-2 hours in 50 mM Na_2CO_3 (with or without 10 mM dithiothreitol) at varying pH from 9.5-12.5. Any insoluble material was removed by centrifugation at 13,000x *g* at room temperature for ten minutes in an Eppendorf minifuge. If this insoluble fraction was to be run on an SDS-PAGE gel then it was washed by resuspension in the same volume of the original buffer and pelleted again.

2.3.7 δ-Endotoxin Protease Treatment

Solublilised inclusion proteins were incubated with trypsin (1:10, w/w, enzyme/toxin). The specific details are given in the text. Prior to SDS-PAGE, the protease was inactivated by addition of PMSF to a final concentration of 2 mM and the sample incubated at 100°C for 5 minutes in gel sample buffer.

2.3.8 N-Terminal Protein Sequencing

Toxin preparations were resolved by SDS-PAGE with the inclusion of 0.015% v/v thioglycolic acid in the upper electrode buffer to scavenge free radicals that can block N-termini and thus impede protein sequencing. Proteins were then transferred to Problot membranes (polyvinylidene difluoride) using an LKB transblot apparatus operating at 3 mA per cm^2 of membrane.

The blotted membrane was rinsed in distilled water for 5 minutes, stained for 3 minutes with 0.2% Coomassie Brilliant Blue R in 50% methanol/10% acetic acid, and then destained in several washes of 50% methanol/10% acetic acid. The membrane was then

briefly rinsed in distilled water, air-dried and then heat sealed in a hybridisation bag before being stored at -20°C until required.

Coomassie Brilliant Blue-stained protein bands were excised from the membrane for determining the N-terminal sequence. This was carried out by Mr. Mike Weldon (Protein Sequencing Facility, Department of Biochemistry, University of Cambridge).

2.3.9 Preparation of Protein Precipitates from Bacterial Cell Cultures

Total protein from sporulating and vegetative Bt cultures was precipitated from 1.5 ml of the bacterial cells in the following manner. Cells (and where applicable spores together with protein inclusions) were recovered by centrifuging at maximum speed for 10 minutes in a minifuge. Lysates were prepared by resuspending pellets in 50 mM Tris-acetate pH 8.0 containing 10% sucrose (w/v) and 100 µg/ml lysozyme to a total volume of 100 µl and incubating for 10 minutes at 37°C. The samples were placed on ice and ice-cold 100 mM Tris-acetate (pH 8.0; 100 µl), 100 mM EDTA (100 ml), 2 mM PMSF/2 mM NEM (100 µl) and 2% (w/v) Triton X-100 (200 µl) was added. The volume was made up to 800 µl by the addition of 50 mM Tris-acetate (pH 8.0)/10% sucrose. Sonication was carried out with bursts of maximum intensity for 30 seconds with a one minute cooling period between bursts (MSE "Soniprep 150", 0.5 cm probe). The extent of cell lysis was monitored microscopically and sonication continued until complete lysis had been achieved. The resulting protein was then concentrated by TCA precipitation (Section 2.3.10).

2.3.10 Trichloroacetic Acid (TCA) Precipitation

Proteins were precipitated by the TCA precipitation method as described by Peterson (1977). Ice-cold TCA was added to a final concentration of 12.5%, thoroughly mixed and placed on ice for 15 minutes. The resulting precipitate recovered by centrifuging in an MSE minifuge at 12,000x *g* for 15 minutes at 4°C and the pellet washed twice in acetone (80% v/v) and dried before being subjected to further analysis.

2.3.11 Mosquito Larvae Bioassays

Bioassays were carried out with 3-4 day old *A. aegypti* larvae. Both rearing and bioassays were carried out at 27°C in a humidified incubator. Batches of about 1-4 thousand larvae, depending upon the success rate of egg hatching were reared in containers 25 x 25 x 15 cm deep with about 5 litres of dH_20 and Farley's Farex baby food as diet. The assays were done in 0.5 ml of dH_2O in an Eppendorf tube, with 5 larvae per tube and a total of 30 larvae per toxin concentration. Negative controls contained dH_2O. Mortality was recorded after 48 hours.

2.4 Photography

2.4.1 Agarose gels

Ethidium bromide stained agarose gels were photographed under ultra-violet light (254 nm) illumination using Polaroid 665 or 667 film and an Olympus OM system with a Kodak red filter.

2.4.2 Phase-Contrast Microscopy

Spore-crystal suspensions were pipetted onto a microscope slide and covered with a coverslip. The suspension was examined by a Nikon phase-contrast microscope under an oil immersion objective (100x) and photographed under a green filter/mono-chromatic interference filter (546 nm) by an Olympus OM system using Polaroid 665 or 667 film.

All other photography was performed by Kim Roswell and Richard Summers.

Three

Chapter 3

Characterisation of 84-I and 17A: Two Related Strains of Bacillus thuringiensis ssp. fukuokaensis

3.1 Introduction

Until the 1980's the discovery of new entomophathogenic strains of Bt had focused on those active against coleopteran and lepidopteran larvae. However, since the discovery of the subspecies *israelensis* (de Barjac, 1978) there has been increasing interest in strains that display toxicity towards dipteran insects.

Control of the larvae of the malaria-carrying mosquito has, for obvious reasons been of primary interest in the search for novel diptericidal toxins. Current biological control of mosquito larvae relies heavily on the use of two entomopathogenic bacteria, *Bacillus thuringiensis* ssp. *israelensis* (Bti) and *Bacillus sphaericus*. Both bacteria synthesise proteins during sporulation that assemble into crystals which are toxic for the larvae upon ingestion. It has generally been accepted that the Bti inclusion contains four major δ-endotoxin proteins of molecular masses 134.4, 127.8, 72.4 and 27.4-kDa deduced from the DNA sequences of their cloned genes. These products are Cry4A, Cry4B, Cry11A and Cyt1A respectively. A fifth gene codes for a 77.8-kDa δ-endotoxin Cry10C, originally thought to be a minor component of the parasporal body (Thorne *et al.*, 1986; Garduno *et al.*, 1988), but which has since been identified as a fifth major crystal component (Purcell, 1997). Crystals from *B. sphaericus* are composed of only two polypeptides of 51 and 42-kDa. The contribution of each of these proteins to the overall toxicity of the host bacteria has been extensively studied. The individual toxins have different specificities and potencies, most of them acting synergistically to increase the overall toxicity of the complete crystal (Angsuthanasombat *et al.*, 1992; Chang *et al.*, 1993; Crickmore *et al.*, 1995; Delécluse *et al.*, 1989; Nicolas *et al.*, 1993; Poncet *et al.*, 1993; Wu *et al.*, 199 4).

Identification of novel mosquitocidal toxins that differ in structure and mode of action from those produced by Bti and *B. sphaericus* may prove to be of great value in improving the entomopathogenic properties of currently used formulations. Combining several genes encoding toxins with different specificity and/or levels of activity in the same organism could lead to the development of new pesticides and may delay or even prevent the onset of resistance. Indeed, there is a report which suggests that, at least in the laboratory, the appearance of mosquitocidal resistance to Bti toxins is inversely correlated

to the number of toxins used for selection (Georghiou *et al.*, 1993 & 1997). Several programs have therefore been established to isolate and characterise mosquitocidally active Bt strains.

There have been several reports of other Bt strains of various serotypes with mosquitocidal activity. These strains are often characterised by the complex polypeptide profile of their crystalline inclusions when subjected to SDS-PAGE and can be classified into four groups according to their larvicidal activity, crystal protein composition and presence of Bti-related genes (Delécluse *et al.*, 1995; Ragni *et al.*, 1996). Group 1 contains strains that display larvicidal activities and crystal polypeptides similar to those of Bti. These include, ssp. *morrisoni* PG14 (H8a8b) (Padua *et al.*, 1984), ssp. *canadensis* (H5a5c) (Ishii & Ohba, 1993), ssp. *thompsoni* B175 (H12) (Brown & Whiteley, 1992) and ssp. *malayensis* IMR81.1 (H36) (Ragni *et al.*, 1996). Group 2 comprises strains that display levels of toxicity close to those of Bti but contain different crystal polypeptides. These include ssp. *medellin* 163-131 (H30) (Orduz *et al.*, 1992) and ssp. *jegathesan* 367 (H28a29c) (Delécluse *et al.*, 1995). Group 3 consists of the strain ssp. *darmastadiensis* 73E10-2 (H10a10b) (Padua *et al.*, 1980), which synthesises different polypeptides compared to Bti and is active against only one mosquito species. Group 4 include the two strains Bt ssp. *fukuokaensis* 84-I-1-13 (Yu *et al.*, 1991) and Bt ssp. *kyushuensis* 74 F6-18 (H11a11c) (Ohba & Aizawa, 1979), which are only weakly active against selected mosquito strains.

Because of the poor toxicity of the group 3 and 4 strains to important disease vectors, they have not been extensively studied. However, despite the low toxicity of the native crystals these strains still present areas of significant interest for study since novel toxins may well differ in structure and mode of action from previously cloned proteins. Also, as discussed above, the discovery of novel toxins increases the options available for combining various toxins in one recombinant cell to prevent or delay the onset of insect resistance.

Yu *et al.* (1991) characterised the strain Bt ssp. *fukuokaensis* 84-I-1-13 (84-I) that was first identified in Japan by Ohba (Ohba & Aizawa, 1989). Crystals from this strain were isolated and used in bioassays against fourth instar larvae of two mosquito species, *Culex quinquefasciatus* and *Aedes aegypti*. LC_{50} values of 4.1 and 2.9 µg/ml respectively were obtained. In addition to this they found that the solubilised crystals displayed haemolytic activity, down to 50 µg/ml. SDS-PAGE analysis revealed that the parasporal crystalline inclusions included a low molecular weight protein of similar size to other known cytolytic toxins in the family of the previously described Cyt toxins (Yu *et al.*, 1991). It was therefore postulated that 84-I contained a novel Cyt toxin that was responsible for the described haemolytic activity. However, recent work carried out by Dunn (1996) in this laboratory found that the reported haemolysis results were not repeatable. Furthermore, subsequent cloning of the 24-kDa protein gene thought most likely to be a Cyt toxin revealed that it was in fact an Orf1 protein. Consequently, none of the mosquitocidal components of this strain

have yet been cloned and characterised. It was therefore decided, for the reasons discussed above, that the identification of potentially novel mosquitocidal toxins from this strain and their subsequent cloning and characterisation would be of value .

In parallel to this study it was also decided to investigate a second serotype of Bt ssp. *fukuokaensis* isolated by Karamanlidou *et al.* (1991) from Greek olive groves (Bt ssp. *fukuokaensis* (H3a:3d:3e) (17A)). The Olive fruit fly, *Dacus oleae*, is a very common insect infesting olive plants through out the Mediterranean basin. Like other species of the Tephritidae family (fruit flies), the pest bores though the pulp causing serious damage to the fruit. Because of the importance of the olive in the economies of several Mediterranean countries, control of this pest is considered vital. However, traditional methods have relied on spraying wide areas with various chemical pesticides which pollute the environment and may affect beneficial species. In the past few years integrated control approaches and products such as food and pheromone traps have been introduced, with various degrees of efficiency, in an attempt to reduce or eliminate the need for such chemical sprays.

The search for further agents of biological control with potential to control the Olive fruit fly prompted the isolation of Bt strains from the environments of Greek olive groves (Karamanlidou *et al.*, 1991), including soil samples from the ground under olive trees and material from tree stalks and local olive-oil producing mills. One strain, 17A (originally called Ormilia after the region in which it was isolated), a new serotype of Bt ssp. *fukuokaensis*, was found to be particularly toxic to *Dacus oleae* larvae with 87% mortality reported within 48 hours.

Since this initial identification of 17A, little further characterisation has been carried out and none of the diptericidal proteins has been identified. In fact, there is no record of any cloned Bt endotoxins that show activity against *D. oleae*. It was therefore decided that the cloning and subsequent characterisation of one or more potentially novel diptericidal proteins from 17A would be of benefit in developing biological insecticides for use against *D. oleae*. In addition, the characterisation of novel toxins active against a new target insect could enhance our knowledge of how these toxins function, particularly in comparison to toxins with different specificity from the related strain 84-I.

3.2 Preparation of δ-Endotoxin Crystals

The preparation of δ-endotoxin crystals from *Bacillus thuringiensis* ssp. *fukuokaensis* 84-I and 17A was carried out essentially as described in Methods (Section 2.3.1). Both strains were streaked out from a frozen glycerol stock onto CCY agar plates and grown at 30°C for two to three days until sporulation had been reached. Individual colonies were then examined by phase-contrast microscopy to ensure that the colonies from which crystals

were to be prepared contained both spores and crystals. Crystals from both strains appeared spherical and phase dark (Figures 3.1a and 3.1b). The selected colonies were then heat-activated to destroy vegetative cells and promote synchronous germination of the spores. The heat-activated spores were inoculated into 200 ml PWYE medium, grown overnight and diluted the next day 1:1000 in CCY medium to encourage sporulation. The culture was then incubated for a further 72 hours until approximately 95% of the spores and crystals had been released, as determined by phase-contrast microscopy. Sporulated cells appeared to have reached the final stage of sporulation (Stage VI, Ryter, 1965) approximately forty-eight hours after inoculation into CCY, it then took an additional twelve hours for the cells to release their spores and crystals. Sodium chloride was then added to a final concentration of 1M to inhibit proteolysis and the culture stored at 4°C for a minimum of one hour (but usually overnight).

Crystals were then separated from spores using a discontinuous sucrose gradient as described in Section 2.3.1. Initial attempts to separate the spores and crystals using a standard gradient of 2.30 M/2.10 M/1.97 M sucrose proved unsuccessful as the crystals appeared to co-migrate with the spores and sediment at the bottom of the tube even with the inclusion of 0.05 M KCl and sonication to reduce clumping. This suggested that not only were the crystals more dense than commonly found but also that they were very similar in density to the spores. A similar problem had been found when workers had tried to separate spores and crystals from Bt ssp. *israelensis*. In this strain it was determined that the crystals existed in two population sizes, only one of which could be separated from the spores (Purcell, 1997). Both 84-I and 17A appear to have several different populations of crystal densities. The most successful sucrose gradient used (2.55 M/2.40 M/2.30 M/1.97 M) gave three populations of crystals. Those in the upper band (between 1.97 M and 2.30 M) had co-migrated with a high concentration of spores as had those in the pellet fraction, while a third population of crystals banded between 2.40 M and 2.50 M and contained only about 10% spores. It was this middle fraction that was used for all further experiments.

Approximately 5, 10 and 15 µg of crystals from each strain were then subjected to SDS-PAGE (Figure 3.3) as described in Section 2.3.4. The gel of 84-I crystal protein revealed nine major protein species ranging in size from 15-kDa to 90-kDa. The eight largest polypeptides (90, 86, 82, 72, 50, 38, 34 and 24-kDa) matched the profile reported by Dunn (1996). A 15-kDa polypeptide which appears to be a major crystal component was not reported by Dunn (1996) although it does seem to be present in previous work carried out by Ishii and Ohba (1993). A similar sized protein from Bt JC292 has recently been cloned by a worker in this lab (D.A. Hinks, Personal communication) and found by immunogold labelling to be a crystal envelope protein. An electron micrograph of the 84-I crystal (Figure 3.2) shows the presence of a similar envelope around the crystal. The 17A crystal protein

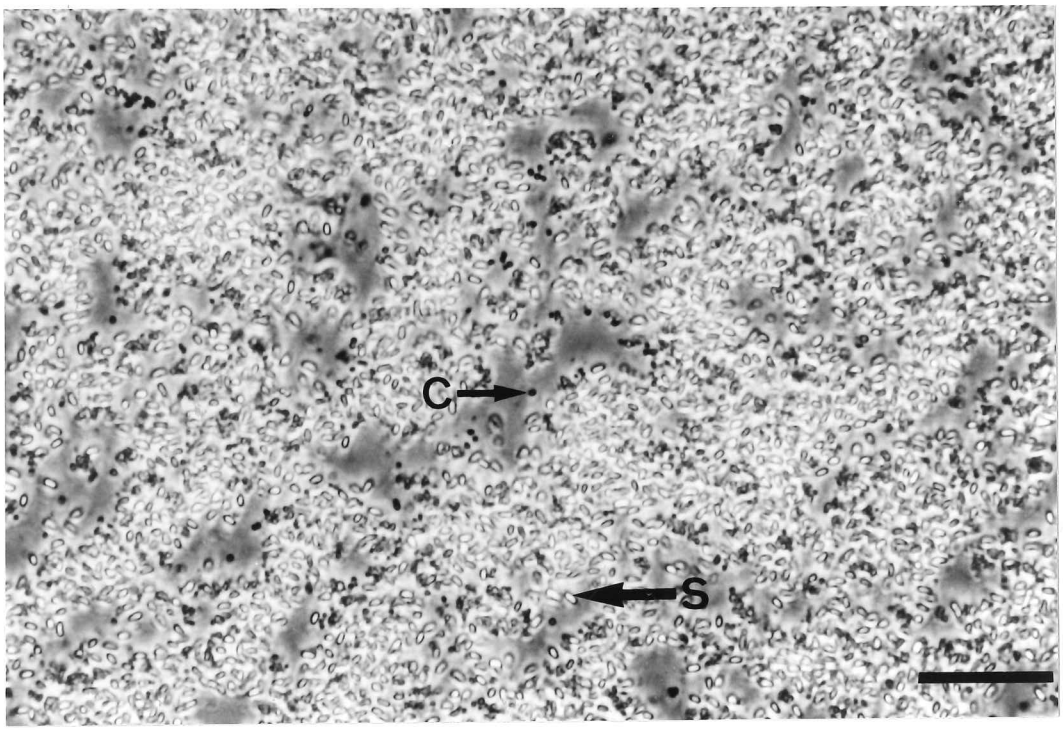


Figure 3.1a Bt ssp. *fukuokaensis* 17A spore/inclusion mixtures as viewed by phase contrast microscopy (photo no. 325E12).
The large arrow (S) indicates a spore and the small arrow (C) indicates a crystalline inclusion. The bar represents 5 µm.

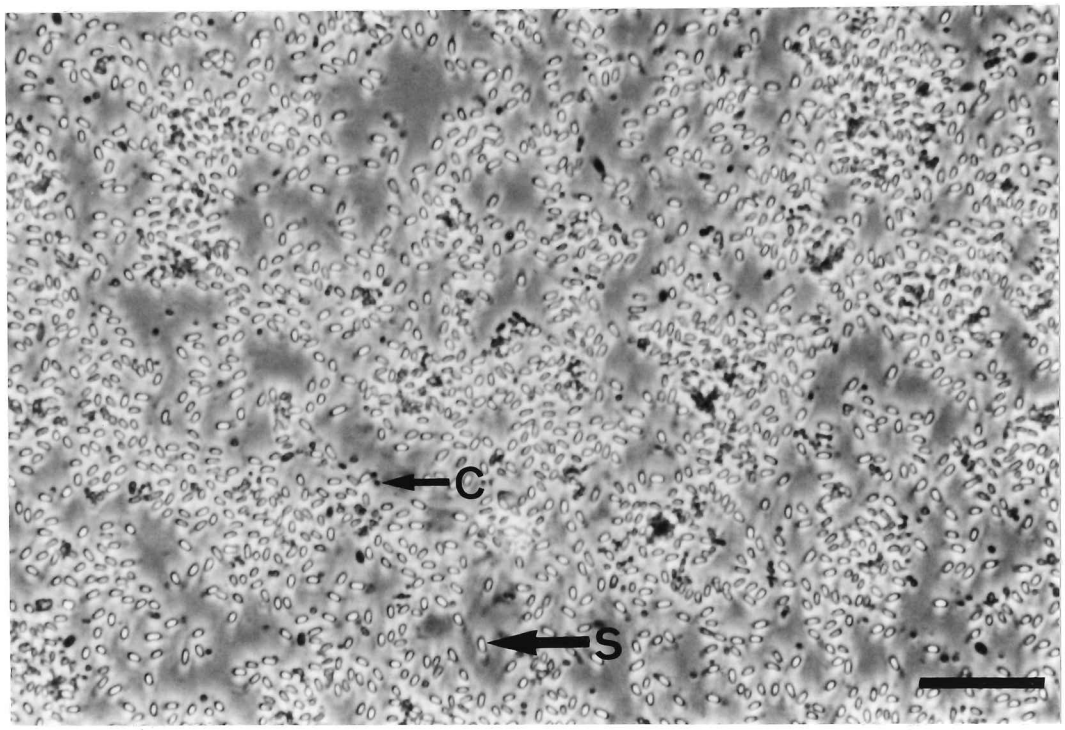


Figure 3.1b Bt ssp. *fukuokaensis* 84-I spore/inclusion mixtures as viewed by phase contrast microscopy (photo no. 325E13).
The large arrow (S) indicates a spore and the small arrow (C) indicates a crystalline inclusion. The bar represents 5 µm.

gel revealed seven major bands at 70, 65, 54, 50, 46, 32 and 15-kDa. Again the presence of a 15-kDa protein may indicate that crystal envelope proteins are a common feature of Bt inclusions. Smearing above the 70-kDa band perhaps indicates that this band is composed of various length degradation products of a larger protein.

Purified inclusions from both strains were therefore found to be composed of a complex set of proteins similar to other Dipteran specific strains. A gel reprinted from Ishii and Ohba (1993) is included for comparison (Figure 3.4). It is interesting to note that both *fukuokaensis* strains investigated here lack the large 130-kDa polypeptides found in the majority of previously identified diptericidal isolates.

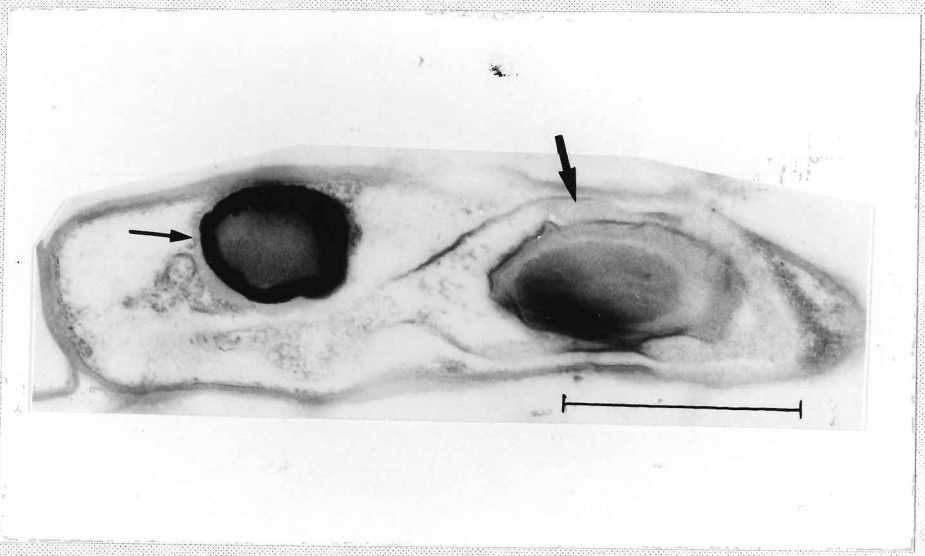


Figure 3.2 Electron micrograph of a stage VI Bt ssp. *fukuokaensis* 84-I cell.
The bar represents 1µm. The spore is indicated by a large arrow and the crystal by a small arrow (photo no. 787D4).

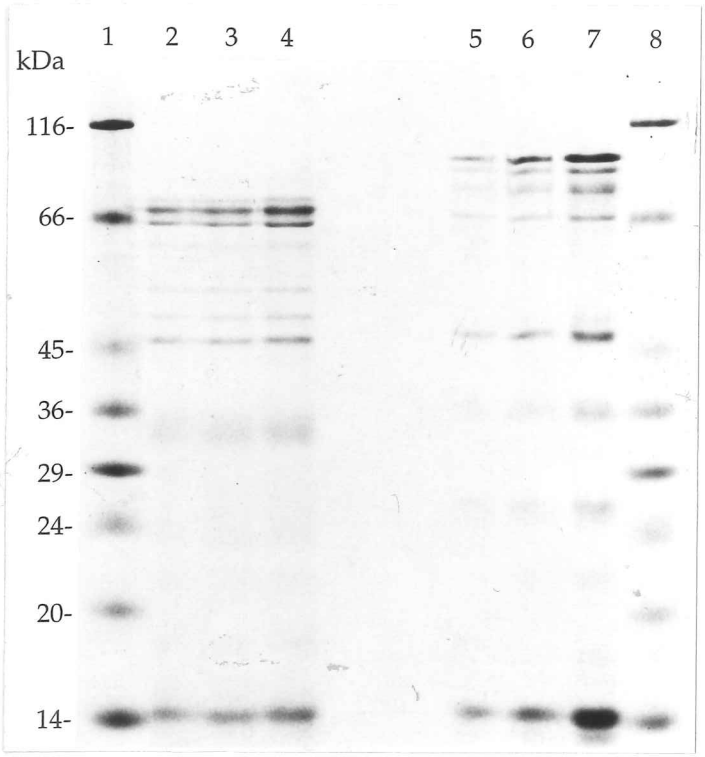


Figure 3.3 Coomassie blue stained 13% SDS/PAGE of sucrose gradient purified crystals from Bt ssp. *fukuokaensis* 17A and 84-I (photo no. 29E2).
Lane 1 Molecular weight markers
Lane 2 5 µg 17A crystals
Lane 3 10 µg 17A crystals
Lane 4 15 µg 17A crystals
Lane 5 5 µg 84-I crystals
Lane 6 10 µg 84-I crystals
Lane 7 15 µg 84-I crystals
Lane 8 Molecular weight markers

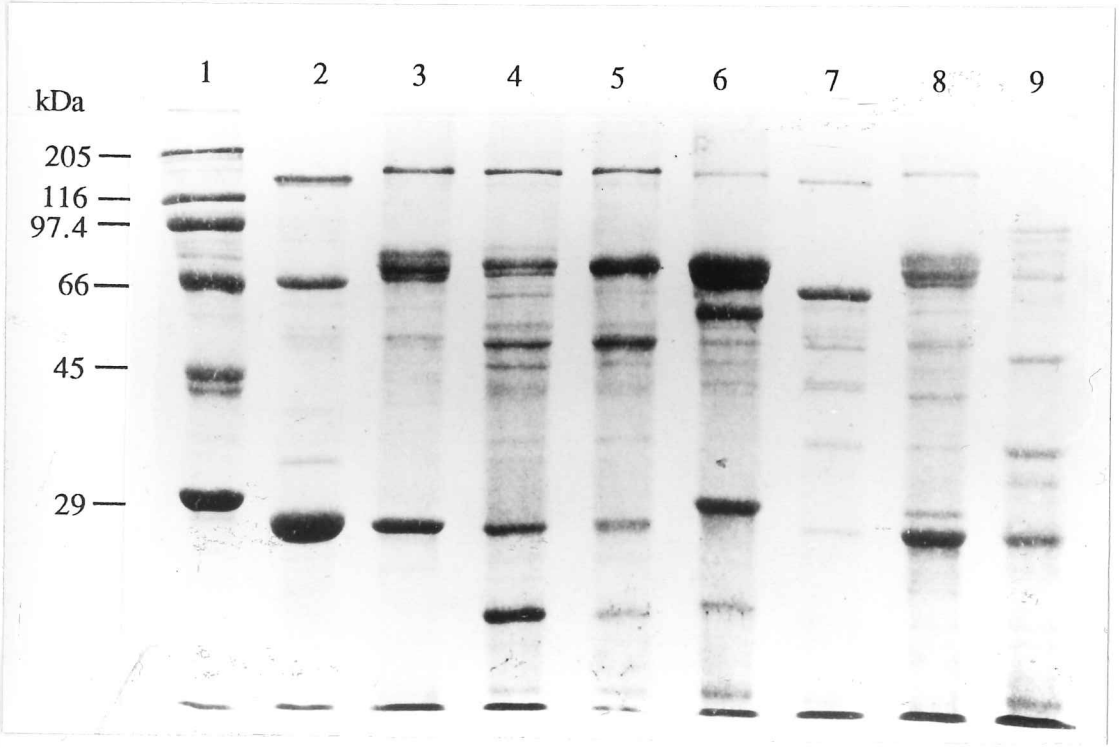


Figure 3.4 Comparative SDS/PAGE of parasporal inclusion proteins from mosquito specific strains of Bt (Reprinted from Ishii & Ohba, 1993) (photo no. 755D3).
Molecular weight markers are indicated on Lane 1
Lane 2 Bt ssp. *israelensis*
Lane 3 Bt ssp. *kyushuensis*
Lane 4 Bt ssp. *darmstadiensis* 73-E-10-2
Lane 5 Bt ssp. *darmstadiensis* 88-KG-2-21
Lane 6 Strain 89-T-34-14
Lane 7 Bt ssp. *candensis*
Lane 8 Bt ssp. *shandongiensis*
Lane 9 Bt ssp. *fukuokaensis*(84-I)

3.3 Solubility and Activation

The process of toxin solubilisation and activation in the insect gut is essential for toxin activity (Jaquet *et al.*, 1987; Murphy *et al.*, 1976) and may also be a factor in determining toxin specificity (Haider *et al.*, 1986). Individual polypeptides from composite crystals can often be solubilised in different conditions to other polypeptides in the same crystal. This feature has sometimes been exploited in order to isolate a single polypeptide away from the composite crystal. The solubility of other dipteran-specific toxins in carbonate/DTT buffers has been studied (Federici *et al.*, 1990) and correlated with a particular toxin type. For example, the 130-kDa Cry4A and Cry4B proteins require alkaline conditions (>9.5) and dithiothreitol for solubilisation (Couche *et al.*, 1987; Federici *et al.*, 1990) whereas the 65-kDa Cry11A protein will solubilise in the absence of reducing agents at pH >10.5. Experiments were undertaken to explore the solubility of the crystal proteins from Bt ssp. *fukuokaensis* 84-I and 17A in the hope that this would indicate the type of δ-endotoxins present in the crystal as well as the best conditions to use for solubilisation in future experiments. 10 µg of sucrose-gradient purified crystals were therefore resuspended in sodium carbonate buffers of varying pH values with or without DTT (Secion 2.3.6).

For 17A, complete solubilisation of the crystals at 37°C was achieved at pH 10 and above in the presence of DTT (Figure 3.5a). The crystal was also completely soluble without a reducing agent at pH 11 and above. Interestingly the 46-kDa protein appears uniquely soluble at pH 9 even in the absence of DTT. There is a degree of proteolysis in all samples so the same assay was carried out at 4°C to minimise degradation by endogenous proteases. Complete solubilisation occurred at pH 10 and above in the presence of DTT, but without the reducing agent the crystal was never completely soluble (Figure 3.5b). At this lower temperature the 46-kDa protein is only soluble below pH 10 in the presence of DTT. However, now that proteolysis has been reduced it can be seen that the 32-kDa and 14-kDa proteins are also at least partially soluble under these conditions.

A study of solubility at 4°C and 37°C leads to the tentative conclusion that putative toxins from this strain solubilise under similar conditions to other known Cry4 δ-endotoxins. It was decided that in future experiments solubilisation of Bt ssp. *fukuokaensis* 17A crystals would be carried out at 4°C using a carbonate buffer (pH 11) containing 10 mM DTT, a regime which involves minimal degradation.

3.3 Solubility and Activation

The process of toxin solubilisation and activation in the insect gut is essential for toxin activity (Jaquet *et al.*, 1987; Murphy *et al.*, 1976) and may also be a factor in determining toxin specificity (Haider *et al.*, 1986). Individual polypeptides from composite crystals can often be solubilised in different conditions to other polypeptides in the same crystal. This feature has sometimes been exploited in order to isolate a single polypeptide away from the composite crystal. The solubility of other dipteran-specific toxins in carbonate/DTT buffers has been studied (Federici *et al.*, 1990) and correlated with a particular toxin type. For example, the 130-kDa Cry4A and Cry4B proteins require alkaline conditions (>9.5) and dithiothreitol for solubilisation (Couche *et al.*, 1987; Federici *et al.*, 1990) whereas the 65-kDa Cry11A protein will solubilise in the absence of reducing agents at pH >10.5. Experiments were undertaken to explore the solubility of the crystal proteins from Bt ssp. *fukuokaensis* 84-I and 17A in the hope that this would indicate the type of δ-endotoxins present in the crystal as well as the best conditions to use for solubilisation in future experiments. 10 µg of sucrose-gradient purified crystals were therefore resuspended in sodium carbonate buffers of varying pH values with or without DTT (Secion 2.3.6).

For 17A, complete solubilisation of the crystals at 37°C was achieved at pH 10 and above in the presence of DTT (Figure 3.5a). The crystal was also completely soluble without a reducing agent at pH 11 and above. Interestingly the 46-kDa protein appears uniquely soluble at pH 9 even in the absence of DTT. There is a degree of proteolysis in all samples so the same assay was carried out at 4°C to minimise degradation by endogenous proteases. Complete solubilisation occurred at pH 10 and above in the presence of DTT, but without the reducing agent the crystal was never completely soluble (Figure 3.5b). At this lower temperature the 46-kDa protein is only soluble below pH 10 in the presence of DTT. However, now that proteolysis has been reduced it can be seen that the 32-kDa and 14-kDa proteins are also at least partially soluble under these conditions.

A study of solubility at 4°C and 37°C leads to the tentative conclusion that putative toxins from this strain solubilise under similar conditions to other known Cry4 δ-endotoxins. It was decided that in future experiments solubilisation of Bt ssp. *fukuokaensis* 17A crystals would be carried out at 4°C using a carbonate buffer (pH 11) containing 10 mM DTT, a regime which involves minimal degradation.

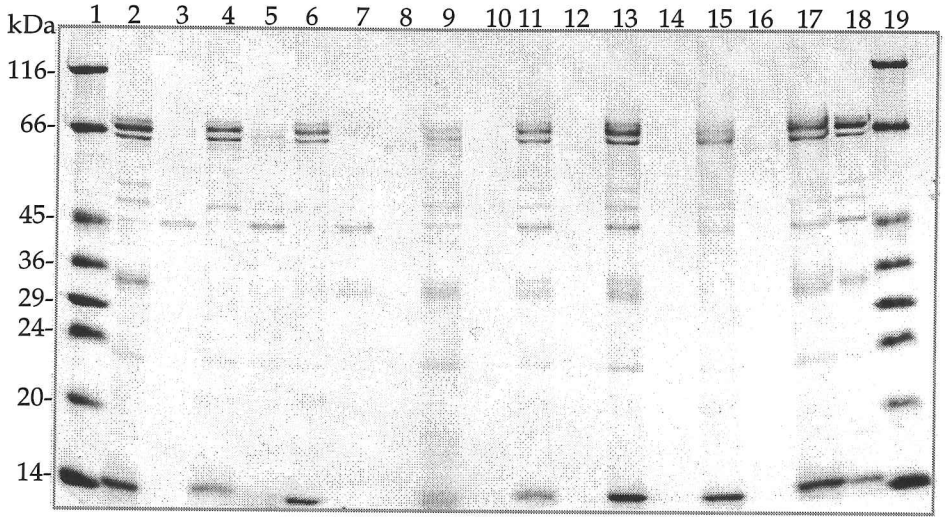


Figure 3.5a Coomassie blue stained 13% SDS-PAGE showing solubility of 10 µg of Bt ssp. *fukuokaensis* 17A crystal proteins in 50 mM Na_2CO_3at varying pH at 37°C, with or without 10 mM DTT (photo no. 935D5).

Lane 1	Molecular weight markers
Lane 2&3	pH 9 -DTT, pellet & supernatant
Lane 4&5	pH 9 +DTT, pellet & supernatant
Lane 6&7	pH 10 -DTT, pellet & supernatant
Lane 8&9	pH 10 +DTT, pellet & supernatant
Lane 10&11	pH 11 -DTT, pellet & supernatant
Lane 12&13	pH 11 +DTT, pellet & supernatant
Lane 14&15	pH 12 -DTT, pellet & supernatant
Lane 16&17	pH 12 +DTT, pellet & supernatant
Lane 18	10 µg 17A crystals
Lane 19	Molecular weight markers

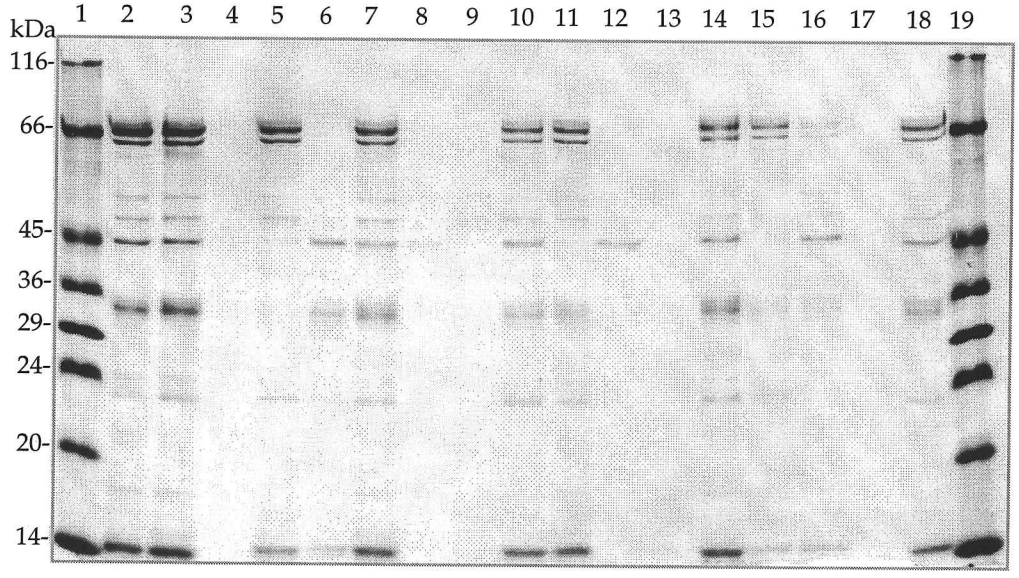


Figure 3.5b Coomassie blue stained 13% SDS-PAGE showing solubility of 10 µg of Bt ssp. *fukuokaensis* 17A crystal proteins in 50 mM Na_2CO_3 at varying pH at 4°C, with or without 10 mM DTT (photo no. 927D5).

Lane 1	Molecular weight markers
Lane 2	10 µg 17A crystals
Lane 3&4	pH 9 -DTT, pellet & supernatant
Lane 5&6	pH 9 +DTT, pellet & supernatant
Lane 7&8	pH 10 -DTT, pellet & supernatant
Lane 9&10	pH 10 +DTT, pellet & supernatant
Lane 11&12	pH 11 -DTT, pellet & supernatant
Lane 13&14	pH 11 +DTT, pellet & supernatant
Lane 15&16	pH 12 -DTT, pellet & supernatant
Lane 17&18	pH 12 +DTT, pellet & supernatant
Lane 19	Molecular weight markers

The solubility of the 84-I crystal was similarly investigated using the same batch of buffers used above. A previous study carried out by Dunn (1996) had found the 84-I inclusion to be completely soluble at pH 10 and above with the addition of DTT and at pH 11 and above without DTT. However, in this instance complete solubilisation was not found to occur at any of the pH values investigated (9-12) even in the presence of DTT (Figure 3.6a). This experiment was repeated several times with freshly made buffers but the result was always the same. Only the lower molecular weight bands (36, 25 and 15-kDa) appeared to be completely soluble, from pH 10 with DTT and at pH 12 without. Solubilisation of the entire crystal was attempted using the harsher conditions of 0.05 M NaOH/10 mM DTT (this solution has a pH of 12), but still only partial solubility was observed (Figure 3.6b).

It was then hypothesised that the discrepancy in 84-I inclusion solubility between that reported by Dunn (1996) and that observed in this study may have been due to the presence of proteases associated with the former crystal preparation. Subsequent to solubilisation in the insect gut the action of gut proteases on the toxins converts inactive protoxins into active toxic fragments that are capable of inserting into the gut epithelial membrane of the insect thus causing cell lysis (Knowles & Ellar, 1987). The processes of solubilisation and activation are thought to occur concomitantly *in vivo* and it was proposed that the presence of proteases may be necessary for efficient solubilisation of 84-I crystals *in vitro*.

To investigate this hypothesis, solubilisation of the 84-I inclusion was carried out in the presence of increasing concentrations of the protease trypsin. Trypsin was chosen as it is the major proteolytic enzyme in the mosquito larval midgut (Borovsky, 1986; Yang & Davies, 1971) against which 84-I has been found to be active. Trypsin was added to 50 mM Na_2CO_3/10 mM DTT at concentrations of 0.04-4 mg/ml (results not shown) and when this was unsuccessful in promoting solubilisation, at 10-100 mg/ml (Figure 3.7). No significant increase in solubility was observed. 84-I crystals were also found to be completely insoluble at acid pH (results not shown).

All the above solubility experiments had been carried out on crystal protein from a single preparation of 84-I inclusions. Separation of crystals and spores using the discontinuous sucrose gradient method previously mentioned was very successful for this particular preparation. Attempts were made to isolate a new batch of 84-I inclusions to determine if the insolubility was a general phenomenon or specific to the first batch. However, repeated attempts to isolate inclusions from several subsequent preparations were totally unsuccessful in that spores and crystals always migrated to the same level (usually both components were found in the pellet fraction). All experiments with varying sucrose densities, spin speeds and spin times were unsuccessful in providing even a relatively pure crystal preparation. The solubility of 84-I inclusions from one of these

spore/crystal mixes was therefore investigated using 50 mM Na_2CO_3/10 mM DTT at pH values of 11 and 12 (Figure 3.8). All crystal components were found to be completely soluble at both pH values (residual insoluble material apparent in the pellet fraction is due to spore proteins).

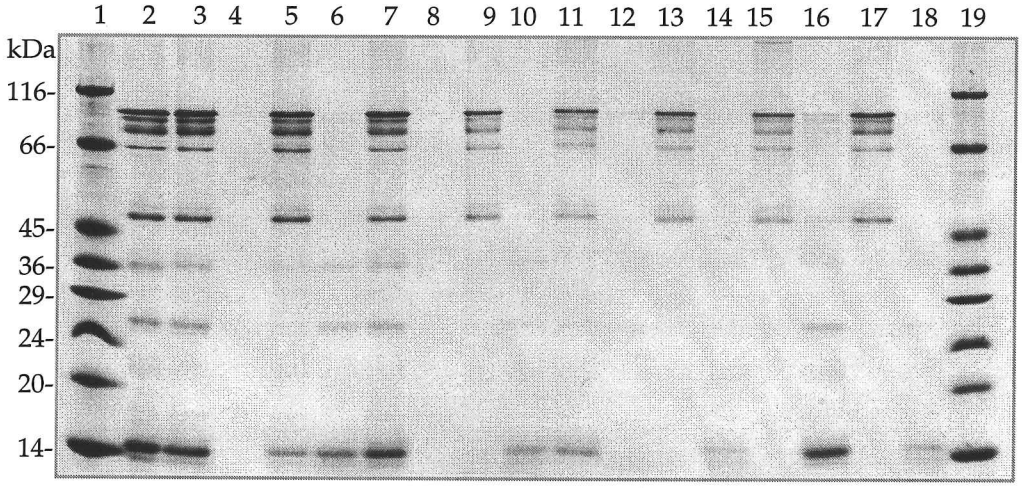


Figure 3.6a Coomassie blue stained 13% SDS-PAGE showing solubility of 10 µg of Bt ssp. *fukuokaensis* 84-I crystal proteins in 50 mM Na_2CO_3 at varying pH at 37°C, with or without 10 mM DTT (photo no. 896D1).

Lane 1 Molecular weight markers
Lane 2 10 µg 84-I crystals
Lane 3&4 pH 9 -DTT, pellet & supernatant
Lane 5&6 pH 9 +DTT, pellet & supernatant
Lane 7&8 pH 10 -DTT, pellet & supernatant
Lane 9&10 pH 10 +DTT, pellet & supernatant
Lane 11&12 pH 11 -DTT, pellet & supernatant
Lane 13&14 pH 11 +DTT, pellet & supernatant
Lane 15&16 pH 12 -DTT, pellet & supernatant
Lane 17&18 pH 12 +DTT, pellet & supernatant
Lane 19 Molecular weight markers

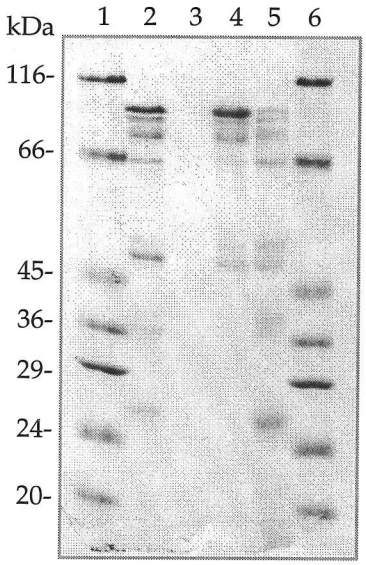


Figure 3.6b Coomassie blue stained 13% SDS-PAGE showing solubility of 10 µg of Bt ssp. *fukuokaensis* 84-I crystal proteins in 0.05 M NaOH/10 mM DTT (photo no. 921D1) at 37°C.

Lane 1 Molecular weight markers
Lane 2 10 µg 84-I crystals
Lane 3 pellet fraction
Lane 4 supernatant fraction

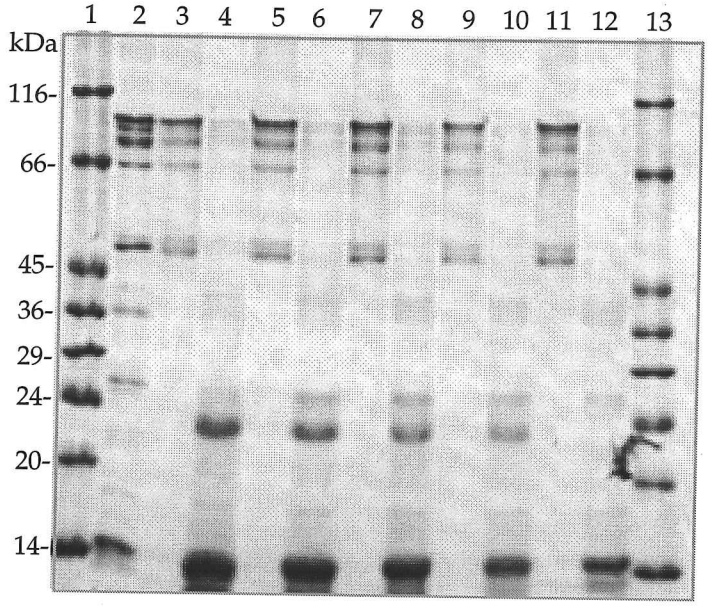


Figure 3.7 Coomassie blue stained 13% SDS-PAGE showing solubility of 10 µg of Bt ssp. *fukuokaensis* 84-I crystal proteins in 50 mM Na_2CO_3/10 mM DTT containing the following concentrations of trypsin (photo no. 951D1) at 37°C.

Lane 1	Molecular weight markers	Lane 7&8	25 mg/ml, pellet & supernatant
Lane 2	10 µg 84-I crystals	Lane 9&10	10 mg/ml, pellet & supernatant
Lane 3&4	100 mg/ml, pellet & supernatant	Lane 11&12	no trypsin, pellet & supernatant
Lane 5&6	50 mg/ml, pellet & supernatant	Lane 13	Molecular weight markers

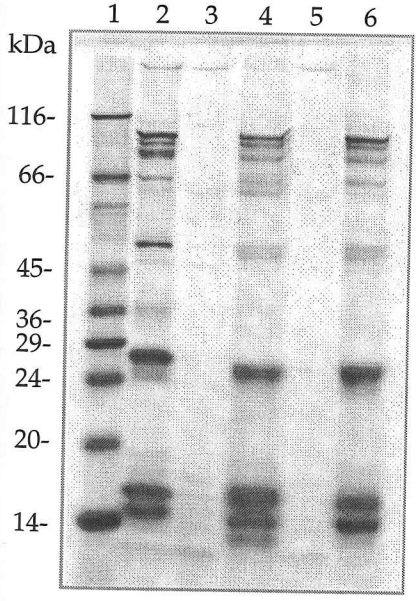


Figure 3.8 Coomassie blue stained 13% SDS-PAGE showing solubility of approximately 20 µg of Bt ssp. *fukuokaensis* 84-I spore/crystal mix in 50 mM Na_2CO_3/10 mM DTT at 37°C (photo no. 958D1).

Lane 1	Molecular weight markers
Lane 2	20 µg 84-I spore/crystal mix
Lane 3&4	pH 11, pellet & supernatant
Lane 5&6	pH 12, pellet & supernatant

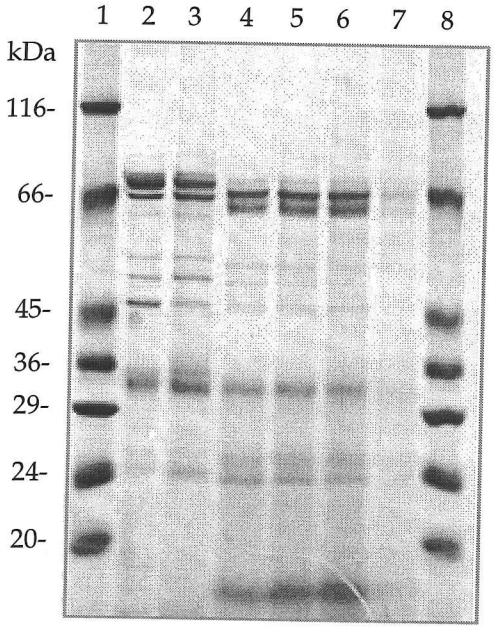


Figure 3.9 Coomassie Blue stained 13% SDS-PAGE gel showing the activation of 17A crystal protein with 5% w/v trypsin (photo no. 958D1) at 37°C.
Lane 1 Molecular weight markers
Lane 2 10 µg 17A crystals
Lane 3 solubilised time zero control
Lane 4 2 hours
Lane 5 4 hours
Lane 6 6 hours
Lane 7 16 hours
Lane 8 Molecular weight markers

No satisfactory conclusions have been reached at this time to suggest why the two preparations described above should show such marked differences in solubility. Suggestions that the insolubility of the first preparation may be to do with very tight packing and perhaps clumping of the crystals would seem to be refuted by the solubility of the lower molecular weight components. The insoluble batch of crystals also appeared to be less dense then their soluble counterparts as they migrated to a lower density sucrose than the spores.

Experiments were carried out to investigate the behaviour of 17A crystal proteins when exposed to proteases. Figure 3.9 shows a 13% SDS-PAGE gel of 10 µg of 17A crystals solubilised at pH 11 as described above and treated for varying lengths of time with 5% w/v trypsin. Major proteins of about 68 and 64-kDa are visible, the size of which compares favourably with the size of most other activated δ–endotoxins. Up until six hours the proteolysed proteins appear to be resistant to further proteolysis, but on extended trypsin treatment very little protein remains intact. The resistance of these proteins to further proteolysis provides evidence that they may be at least partly responsible for the toxic activities of the strain.

3.4 N-terminal Sequence Analysis

Previous work by Dunn (1996) identified the N-terminal amino acid sequence of four of the proteins from 84-I (the 24, 32, 50 and 72-kDa polypeptides). To obtain further information on the proteins present in the crystals of Bt ssp. *fukuokaensis* 84-I and 17A, N-terminal sequencing was carried out on several bands from both profiles as described in Section 2.3.8. The only additional protein from 84-I to yield unambiguous N-terminal sequence data was the 90-kDa polypeptide. Many of the bands from the 17A profile generated mixed sequences despite appearing as clear discrete bands on the blot. Unambiguous sequence was obtained for the 65-kDa polypeptide although the identity of the third residue was unclear. The smeared 70-kDa band produced mixed sequence data as might have been predicted, however it was possible to identify a very similar N-terminal sequence to that of the 65-kDa polypeptide within this data.

N-terminal amino acids sequence analysis was also carried out on the 68 and 64-kDa bands from the activated 17A crystal and, while the lower band gave mixed sequence, nine residues of clear sequence was obtained for the 68-kDa band. This sequence is included in the following table along with the unambiguous N-termini obtained to date for 84-I and 17A.

Protein	Source	N-termini
24-kDa (Dunn)	84-I	MENRNANQHSDYPNHNEYE
32-kDa (Dunn)	84-I	MQKRQLLNLHEGDDIYV
50-kDa (Dunn)	84-I	MNRKQQRFXDSN
72-kDa (Dunn)	84-I	MNPYQNKNEYEFN
90-kDa	84-I	MNPYQNNDEIVD
65-kDa	17A	MN?YQKKNE
68-kDa (Activated)	17A	YPLANDPQA

Comparison of the above sequences with the N-terminal amino acid sequence of previously identified Cry toxins, revealed that only the 72, 90 and 65-kDa proteins showed any significant level of similarity to known Cry proteins. All three of these very similar N-termini showed significant homology to the diptericidal Cry proteins from *Bacillus thuringiensis* ssp. *israelensis* (Bti). An alignment is shown in Figure 3.10.

Cry4A (Bti)	**MNPYQ**NKN**E**YEIFNA
Cry4B (Bti)	**MNPYQ**NKN**E**YETLNA
Cry10A (Bti)	**MNPYQ**NKN**E**YEIFNA
Cry11A (Bti)	**MNPYQ**NKN**E**YETLNA
72-kDa (84-I)	**MNPYQ**NKN**E**YEFN
90-kDa (84-I)	**MNPYQ**NND**E**IVD
65-kDa (17A)	**MN?YQ**KKN**E**
	***** *

Figure 3.10 Alignment of the N-termini of the Cry proteins from the diptericidal Bti with the 72 and 90-kDa proteins from 84-I and the 65-kDa protein from 17A.
Identical residues are in bold and marked with an asterisk.

The 72-kDa protein from 84-I is identical in its first ten residues to the Bti proteins. The first six residues of the 90-kDa protein and possibly (dependent upon the ambiguous third residue) the first five of the 65-kDa are also identical to the Bti proteins and to each other. The similarity of the three *fukuokaensis* N-termini to those of the known dipteran active Bti components suggests that these three proteins may contribute to the diptericidal activity of the two *fukuokaensis* strains.

Sequence similarity searches were carried out on the remaining three N-terminal sequences using BLAST from the GCG package of programs (Section 2.1.6). No significant similarity was found to any previously identified proteins.

3.5 Immunoblot Analysis of Protein Homology

N-terminal sequence analysis had shown that the higher molecular weight proteins in each strain were likely to be related to the Cry proteins of Bti. However, only a very limited number of residues can be compared in this manner and the results do not always fully indicate true relationships between proteins. For instance Cry10A and Cry11A have identical N-terminal sequences down to the eleventh residue; they are also of similar size (about 70-kDa). However, when these two proteins are aligned along their whole length they show less than 15% similarity and are immunologically distinct. The immunological cross-reactivity between the proteins in the crystals of the two Bt ssp. *fukuokaensis* strains and the diptericidal proteins of Bti was therefore investigated by Immunoblot analysis (Section 2.3.5) to gain further insight into the type of δ-endotoxins present in the crystal and aid selection of potentially novel polypeptides for future cloning experiments.

10 µg of Bt ssp. *israelensis* crystals, 10 µg of 17A crystals and 10 µg of 84-I crystals were subject to SDS-PAGE. Individual blots from these gels were then probed with antisera

to the four Bti Cry proteins using a blocking system of 3% w/v haemoglobin in TBS (Figure 3.11). There appeared to be no cross-reactivity between Cry4A antiserum and either Bt ssp. *fukuokaensis* strain but faint cross-reactivity was observed with Cry4B antiserum to proteins of approximately 70-kDa in both strains and also for the 50-kDa protein of 84-I.

Probing with Cry10A antiserum gave the most marked cross-reactivity for both strains. The 70-kDa protein of 17A gives a very strong clear band indicating strong similarity with Cry10A, or at least with the particular antigenic determinants that the polyclonal Cry10A antiserum recognise. Cry10A antiserum also cross-reacts markedly with a similar sized protein (66-kDa) in 84-I. The similarity in size between these proteins and the cross-reactivity with Cry10A antiserum strongly suggests that they are Cry10A-type proteins. The slight difference in size may indicate that they are novel proteins but it may also just be an indication of different endogenous proteolysis occurring to the same protein in the three strains.

Cry10A is considered to be a natural C-terminal truncate (Thorne *et al.*, 1986). Previous studies of Cry10A (Purcell, 1997) indicated that although it may be synthesised as a protein greater than 70-kDa the lack of a protease-resistant C-terminal region causes it to be quickly degraded down to the first protease-resistant site (65-kDa in the case of Bt ssp. *israelensis* although it runs lower in the total Bt. ssp. *israelensis* preparation, pushed down by the Cry11A protein that runs directly above it). Cry10A antiserum also strongly cross-reacts with the 90-kDa protein from 84-I. If this is indeed a Cry10A-type protein then at 90-kDa it will be a novel type. Fainter cross-reactivities with Cry10A antiserum are also obvious in 84-I for a large range of lower molecular weight proteins, perhaps indicating that they are degradation products of the 90 and 66-kDa proteins. The 65-kDa protein of 17A also has a very faint cross-reaction with this antiserum; again it may be a degradation product but the faintness of the reaction suggests that this is another distinct protein that shares some similarity to Cry10A. A moderately strong cross-reactivity to Cry10A antisera is evident at about 90-kDa for 17A even though a protein band is not evident on the Coomassie blue stained gel. This may be evidence of the existence of a small amount of a full length product most of which is endogenously degraded to 70-kDa as is thought to be the case for Cry10A.

Interestingly, this "invisible" 90-kDa band from 17A also cross-reacts with Cry10A antisera, as does a 130-kDa band from the same strain. The identification of crystal components by immunoblotting that are poorly visualised by SDS-PAGE is not unusual. Kawalek *et al.* (1995) reported the presence of two proteins at 105 and 95-kDa in the diptericidal strain Bt ssp. *jegathesan* that cross-reacted with antisera raised against the whole parasporal inclusion of Bt ssp. *israelensis*. These two proteins were not obvious in the SDS-PAGE profile of Bt ssp. *jegathesan* (Figure 3.12). Intriguingly, this strain has a very similar polypeptide profile to that of 17A with at least ten polypeptides reported with

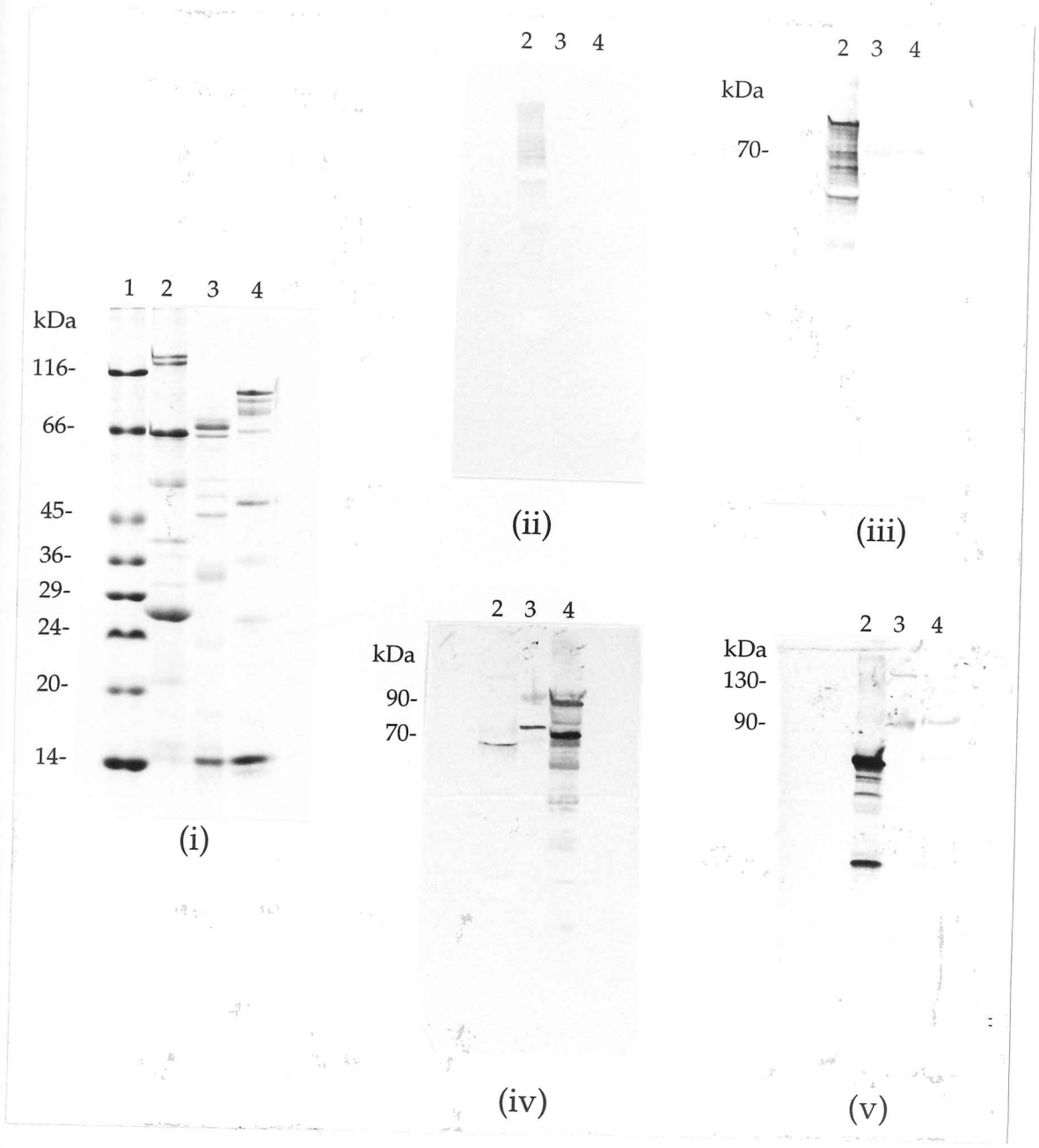


Figure 3.11 Immunoblot analysis of Bt ssp. *fukuokaensis* 17A and 84-I with antibodies raised against the Cry proteins of Bti.

Panel (i) Coomassie Blue stained 13% SDS-PAGE loaded as follows:-
Lane 1 Molecular weight marker
Lane 2 10 µg Bti crystals
Lane 3 10 µg 17A crystals
Lane 4 10 µg 84-I crystals

Panel (ii) Immunoblot of gel as in Panel (i) using Cry4A antiserum.

Panel (iii) Immunoblot of gel as in Panel (i) using Cry4B antiserum.

Panel (iv) Immunoblot of gel as in Panel (i) using Cry10A antiserum.

Panel (v) Immunoblot of gel as in Panel (i) using Cry11A antiserum.

molecular masses of 77 to 23-kDa. The discovery of two higher molecular weight bands in 17A increases the similarity of the two polypeptide profiles. Bt ssp. *jegathesan* was isolated in Malaysia and inclusions from this strain found to have comparable mosquitocidal activity to Bt ssp. *israelensis*. The toxicity of 17A inclusions to mosquito larvae had not been tested but its reaction to antisera against Bti proteins and the similarity of the polypeptide profile to that of Bt ssp. *jegathesan* suggested that these assays should be performed (Section 3.7).

The 65-kDa protein of 17A also cross-reacts very faintly with the Cry11A antiserum as it did with Cry10A, again indicating that this is a novel Cry protein. Similarly in 84-I, the 90-kDa polypeptide shows the strongest cross-reaction with to Cry11A antiserum. Weaker cross-reactions with the 66-kDa and 25-kDa proteins provide further evidence (along with the reaction to Cry10A antisera) that these bands are degradation products of the 90-kDa protein, perhaps created by cleavage at a single site to yield the two unequal halves. Cross-reaction with a 130-kDa band is also evident in 84-I as it was in 17A. The existence of an approximately 130-kDa protein in both Bt ssp. *fukuokaensis* strains is interesting as the 125-kDa Cry4A and 130-kDa Cry4B from Bt ssp. *israelensis* are both major diptericidal toxins. The 130-kDa proteins from 84-I and 17A do not however cross-react with antiserum raised to either Cry4A or Cry4B indicating that the two Bt ssp. *fukuokaensis* strains may contain genes for novel 130-kDa diptericidal proteins that are either very poorly expressed or are degraded by endogenous proteinases on synthesis. The multiple bands seen at 70-kDa on an SDS-PAGE of 17A crystals (Figure 3.3) may be the protease resistant core of a 130-kDa protein from 17A.

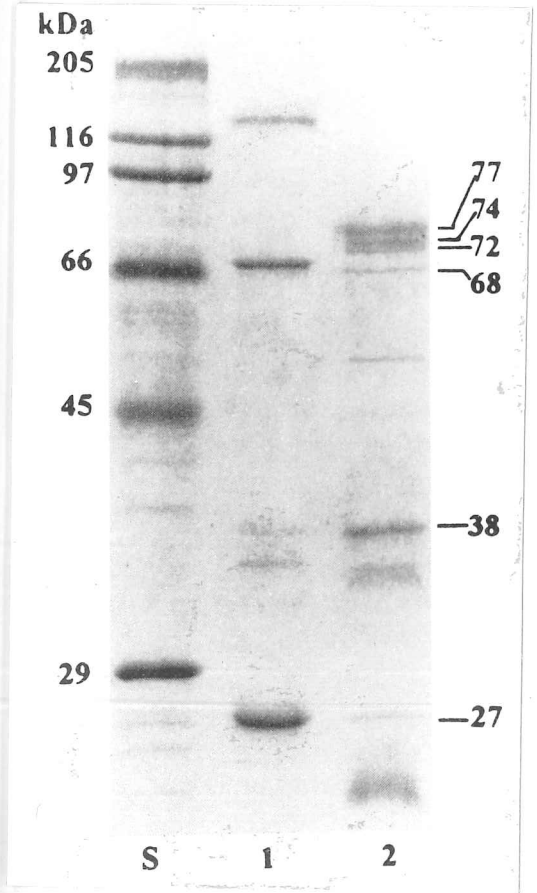


Figure 3.12 Coomassie Blue stained SDS-PAGE.
Crystal protein from Bti (lane 1) and Bt ssp. *jegathesan* (lane 2). Molecular weight markers are in Lane S. Taken from Kawalek *et al.* (1995).

As discussed in the introduction to this Chapter, Bti contains a fifth diptericidal protein, Cyt1A. To investigate the presence of possible Cyt toxins in the two *fukuokaensis* strains, 10 µg of crystals from each strain was subjected to SDS-PAGE alongside either 10 µg of crystals from Bt ssp. *israelensis* (containing Cyt1A) or 1 µg of cloned Cyt1B expressed in *E. coli* (Figure 3.13). Cyt1A antiserum cross-reacted with some of the higher molecular weight proteins from 17A and 84-I as it does with similarly sized proteins from Bti. But no cross-reaction was observed for any of the lower molecular weight proteins that might have been Cyt toxins. Nor was there any cross-reaction with Cyt1B antiserum to any protein present in 17A or 84-I. Although it has been shown that Cyt1A and Cyt1B are 39% identical at the protein sequence level (Koni & Ellar, 1993) they do not exhibit cross-reacting epitopes (Figure 3.13). Thus the lack of immuno-cross reactivity shown here does not exclude the presence of a similar protein in 17A or 84-I. Initially the 24-kDa protein was thought to be the most likely candidate for a Cyt protein from 84-I as it is a similar size to other Cyt toxins and was reportedly haemolytic (Yu *et al.*, 1991; Ishii & Ohba, 1993). However, Dunn (1996) found that the reported haemolysis results were not repeatable despite the fact that the crystal polypeptide profiles obtained in his work and that of Ishii and Ohba (1993) were identical. It is therefore possible that no such cytolytic protein exists in 84-I.

The 24-kDa protein was cloned by Dunn (1996) and found to be an Orf1 protein. The function of the Orf1 family of genes is unknown, but they have been suggested to have a role in crystal formation and or/stability (Dervyn *et al.*, 1995). The existence of *orf1* genes has been correlated with strains where *cry2* genes are expressed. Cry2 proteins are naturally truncated and lack the cysteine-rich C-terminus, a region which has been suggested to be important in crystal packing and stability (Höfte & Whiteley, 1989). While *orf1* correlates with the presence of *cry2* genes, *p22* and *p19* (two other genes found associated with δ-endotoxin producing operons) both appear to be present in strains which have truncated *cry4*-type genes. *p19* is located upstream of *cry11A* (Dervyn *et al.*, 1995), a 70-kDa protein with a Cry4-like N-terminus (see Figure 3.10). The existence of Orf1 in 84-I may therefore be correlated with the presence of truncated Cry4-type proteins as indicated by the above immunoblotting experiments. Likewise, 17A has been shown to contain truncated Cry4-type proteins (see Figure 3.10). 17A also possesses two similarly sized proteins to the 84-I Orf1 and it was thought possible that one or both of these could be Orf1 proteins. Antiserum raised against the Orf1 protein from 84-I was used to probe a further blot of crystal protein from Bt ssp. *israelensis*, 17A and 84-I (Figure 3.14). As can be seen from the blot, no cross-reaction is evident with any of the 17A or *israelensis* proteins indicating that if the 17A crystal does contain an Orf1 protein it is not immunologically related to the Orf1 from 84-I.

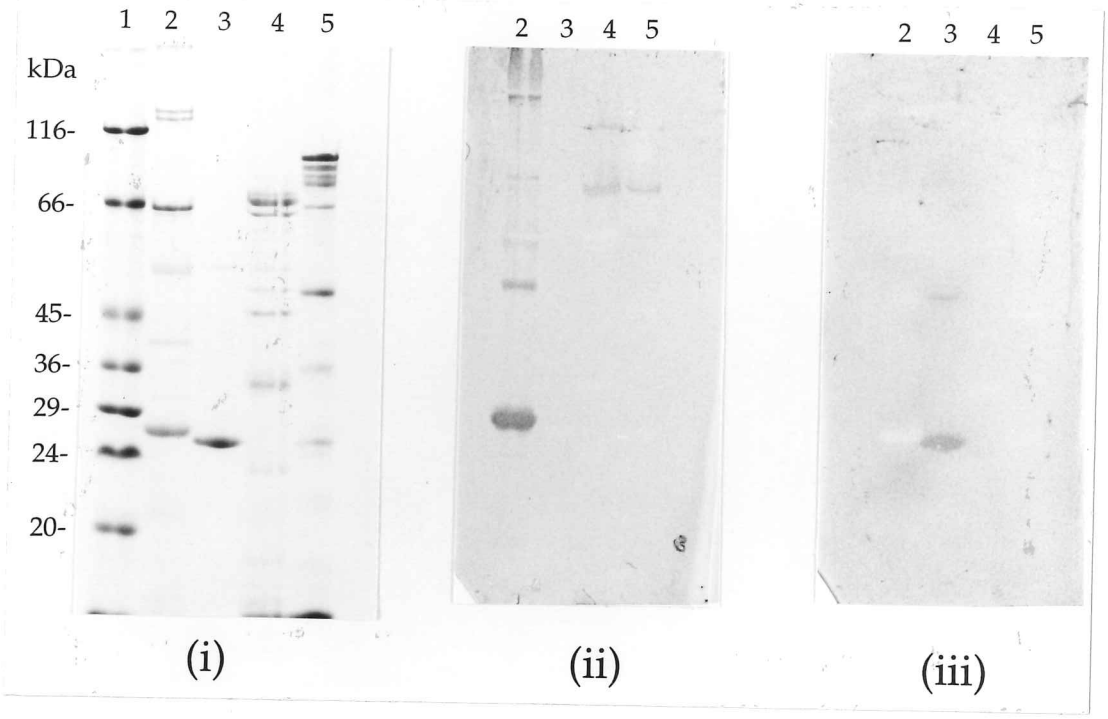


Figure 3.13 Immunoblot analysis of and Bt ssp. *fukuokaensis* 17A and 84-I with antibodies raised against Cyt1A and Cyt1B.

Panel (i) Coomassie Blue stained 13% SDS-PAGE loaded as follows:-
Lane 1 Molecular weight marker
Lane 2 10 µg Bti crystals
Lane 3 1 µg cloned Cyt1B crystals
Lane 4 10 µg 17A crystals
Lane 5 10 µg 84-I crystals

Panel (ii) Immunoblot of gel as in Panel (i) using Cyt1A antiserum.

Panel (iii) Immunoblot of gel as in Panel (i) using Cyt1B antiserum.

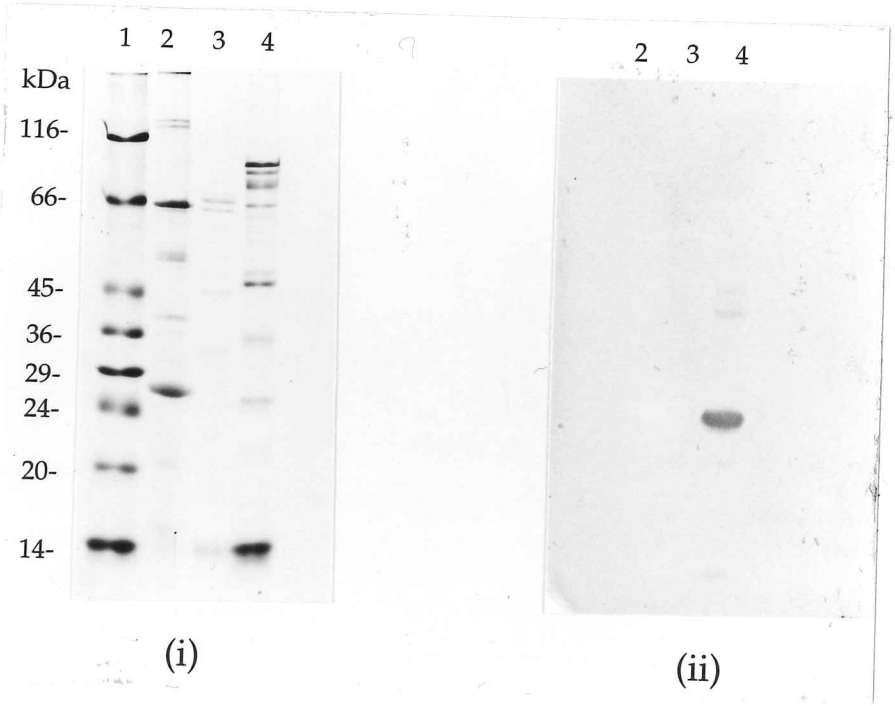


Figure 3.14 Immunoblot analysis of and Bt ssp. *fukuokaensis* 17A and 84-I with antibodies raised against the Orf1 protein of 84-I.

Panel (i) Coomassie Blue stained 13% SDS-PAGE loaded as follows:-
Lane 1 Molecular weight marker
Lane 2 10 µg Bti crystals
Lane 3 10 µg 17A crystals
Lane 4 10 µg 84-I crystals

Panel (ii) Immunoblot of gel as in Panel (i) using Orf1 antiserum.

3.6 PCR Screening for Protein Homology

In recent years, Polymerase Chain Reaction (PCR) with primers specific to the different *cry* genes has allowed rapid and highly sensitive determination of the *cry* gene content of a number of *Bacillus thuringiensis* strains (Ceron *et al.*, 1995; S.P. Gash, personal communication). In this work, PCR was used in an attempt to identify any previously cloned genes for known diptericidal proteins that might also be present in 17A or 84-I. The PCR primers used in this work were designed by S.P. Gash (personal communication) to screen multiple Bt strains. The primer pairs were designed using the PrimerSelect software of the DNASTAR Lasergene package. This program finds the best primer pairs within a sequence of DNA with respect to suitable annealing temperatures, primer stability, avoidance of primer dimers and hairpins. Primers were selected within regions of low homology to other known *cry* genes and have been shown to be specific for the gene to which they were designed (S.P. Gash, personal communication).

Primers were selected to screen for the presence of fifteen known diptericidal genes. These are listed alongside the primer pairs in Figure 3.15. Screening for *cry2* type genes was of particular interest as their existence in 84-I had been implied by the discovery of *orf1* as discussed in the above immunoblotting experiments.

Crude preparations of DNA were extracted from 17A and 84-I as described in Section 2.2.14 and also from strains containing the genes listed in Figure 3.15 to provide positive controls. The *fukuokaensis* strains were then screened for the presence of the above genes using the following PCR conditions;

Denaturation	94°C	1 minute
Annealing	see table	1 minute
Extension	72°C	1 minute

20 µl of each PCR reaction was run out on a 1% agarose gel in 1x TBE together with a 100 bp ladder. A product of the correct size was identified for each of the positive controls but no products were obtained for any of the primer pairs when 17A or 84-I DNA was used as the template (results not shown). From this it can be concluded that none of above genes are present in either 17A or 84-I. Although it is possible that other known genes that were not tested for in this round of screening are present in 17A and 84-I it does appear likely that both strains contain at least some novel genes. Of particular interest is the fact that no product was obtained using the *cry10A* or *cry11A* primer pairs. This indicates that the polypeptides of around 70-kDa in both strains that cross-reacted with Cry10A and Cry11A antisera are not identical to either of these proteins although they are likely to share at least some sequence identity.

Gene	Reference	Upper Primer 5'-->3' Lower Primer 5'-->3'	Optimal Annealing Temp.
cry4A	Ward & Ellar, 1987	CCAAACCCACAAAATACTCAGGATG GAATCAAGCCAAGTAAATAAATGCGGTCTA	53.0°C
cry4B	Yamamoto *et al.*, 1988	ATACCCTGCGGACAAAATAGATAATACGAAACT TTATAAGAGGCAAGAGTACCATCAATTTTGTAGA	53.3°C
cry10A	Thorne *et al.*, 1986	TATTGTTGGAGTTAGTGCAGGTATTATTGTAG TATTCCATGTTGCGTTAGTATTAGTTC	51.8°C
cry11A	Donovan *et al.*, 1988	TAGGGGGATTATTGAGGTTAGTGATGTATTTGA TCCACCCATATAGAGTAGACTGAAACCAC	58.2°C
cry2Aa	Donovan *et al.*, 1988	GGATATTGAGTGAATTATGGGGATA CCGCTATAATTAACCCTGGCACTATTCAATGA	52.8°C
cry2Ab	Widner & Whiteley, 1989	CACAGCAGACCCAATCATTTACTTCACAAGA CTGTAAAAGCACCACTCCTTAACCCTAAA	55.1°C
cry2Ac	Wu *et al.*, 1991	GGAGTGTCATCTAGCCCGCATAGGTCAAG ACCATAATATTCATAAGCTCAAATTGTGGATTGC	52.7°C
cry1Ba	Brizzard & Whiteley, 1988	ATGCAAGAACGAGAAGTGTTCTTTATACC CCTATTGGCCGAGATTGAATCGTGTGC	55.2°C
cry1Bb	Donovan *et al.*, 1994	TTATTTAATTCGTTACAATGCAAAG TTACGGGAATTACATGCTGCTGT	53.5°C
cry1Bc	Bishop, 1994 (unpublished)	GAAGAAGTGGAGGGATAAGAGAGAGAAGTTACAC TTCATTACGGGAACCACATTCATTAC	54.2°C
cry1Ca	Honee *et al.*, 1988	ATTGGGGAGGACATCGAGTAATATCTAGCCTTAT ACTCCTGTGGATGCCGCTCCTGTTAATACTAT	57.0°C
cry1Cb	Kalman *et al.*, 1993	GGACTTTATCAAATCCTACTTTTAGACCTT TCCTCAAATGTAGAAGTAACTGGGAGAA	54.3°C
cry1J	Donovan *et al.*, 1994	TCAGATCCTGATAACGAAGCGGCTAAAA GTCGCCTTATCTGCTGTTCAGTAAGAGTACCA	55.1°C
cyt1A	Waalwijk *et al.*, 1985	GATCAAAATGTAACTCAAACGAATAACCAA TACACAATACATAACGCCACCAGT	50.5°C
cyt1B	Koni & Ellar 1993	AGAATGGAAGTAAAAGGTAATAACGGGTGTTC TGCTAAAGGTAATACTGCCATAACTGC	51.8°C

Figure 3.15 Primers used to identify the presence of any known diptericidal genes present in Bt ssp. *fukuokaensis* 17A and 84-I (S.P. Gash personal communication).

3.7 *In vivo* Toxicity Assays

The toxicity of Bt ssp. *fukuokaensis* 84-I δ-endotoxin inclusions towards three to four day old *Aedes aegypti* mosquito larvae had been tested by Dunn (1996). The LC_{50} value was determined to be 3.5 µg/ml, this is in close agreement to the value obtained by Yu *et al.* (1991). Bioassays were also carried out on the Mushroom fly *Lycoriella mali* by Clifford Keil of the University of Delaware (Dunn, 1996). Even at very high concentrations (25 mg/ml of toxin) little toxicity was seen.

As discussed above, Bt ssp. *fukuokaensis* 17A was originally isolated as being toxic to the Olive fruit fly (*Dacus oleae*). In the original study by Karamanlidou *et al.* (1991) toxicity assays were performed on five day old larvae and one to seven day old flies. 87% mortality was recorded towards larvae and 15% towards flies after 48 hours. These values were higher than obtained for any other isolate. No LC_{50} value was obtained in this study although it was noted that when purified crystals and spores were combined they always induced higher levels of mortality than the same quantity of either crystals or spores.

The mosquitocidal activity of the related strain 84-I, the similarity in polypeptide profile to the highly mosquitocidal strain Bt ssp. *fukuokaensis* and the identification of Cry proteins similar to those in the mosquitocidal Bti, all indicated that 17A may also possess mosquitocidal activity. To test this theory, the toxicity of 17A spore/inclusion mix towards three and four day old *Aedes aegypti* mosquito larvae was assayed as described in Section 2.3.11. Five mosquito larvae were used for each concentration of toxin and each concentration was repeated six times. It was found that spore/crystal mixes containing up to 300 µg/ml of 17A crystal were non-toxic to these mosquito larvae.

3.8 Discussion

In this Chapter, the polypeptide profile of Bt ssp. *fukuokaensis* 84-I and the related strain 17A were investigated with a view to identifying suitable proteins for subsequent cloning experiments.

In order to analyse the polypeptides that comprise the crystals of both strains, spores were removed using discontinuous sucrose gradients. Difficulties were found in achieving satisfactory separation of the crystals and spores from both strains; this was thought to be due to the spores and crystals being similar in size and density. Good separation was eventually obtained although it should be noted that even using the final conditions stated in this Chapter, isolation of pure crystal protein was not always possible. Different preparations seem to vary, perhaps indicating that slight differences in growth conditions (e.g. temperature, aeration) may effect the size and density of the crystal formed

and thus the ease of isolation. For 84-I, inclusion solubility also seemed to be affected. Purified inclusions from both strains were found to be composed of a complex set of proteins similar to other Dipteran-specific strains but lacking the large 130-kDa proteins found in the majority of previously identified isolates.

In view of the difficulty and time consuming nature of purifying 17A and 84-I inclusions on a sucrose density gradient, solubilisation of the spore/crystal mix and subsequent use of the supernatant fraction may provide a more practical method of obtaining pure crystal protein from these strains.

N-terminal amino acid, immunological and PCR analysis indicated that both strains are comprised of a completely novel set of δ-endotoxin and other proteins, although some similarities to previously identified Cry proteins were noted.

N-terminal amino acid sequence analysis indicated that the larger proteins in each profile (90 and 72-kDa of 84-I, and 70 and 65-kDa of 17A) were related to the Cry proteins of Bti. Immunoblotting experiments confirmed that Cry10A-type proteins were present in both strains, although subsequent PCR carried out using primers specific to Cry10A did not give a positive reaction for either strain indicating that the Cry10A-types were indeed novel. This initial exploration suggests that both strains contain at least two novel Cry10A-type proteins.

In identifying genes for subsequent cloning, the 90-kDa protein of 84-I is of particular interest. The above investigation strongly indicates that not only is it responsible for at least part of the mosquitocidal activity of 84-I (as implied by the similarity of its N-terminus to Cry proteins from the mosquitocidal Bti) but also that it is a novel protein. At 90-kDa it would be a novel size for a Cry10A-type protein, the N-terminus diverges from the consensus Bti Cry protein termini after only six residues and none of the immunoblot cross-reactivities are as strong as the control. It is also a major component of the 84-I crystal as shown by the density of the band visible on SDS-PAGE. The 72-kDa protein on the other hand shows much stronger similarity in size, N-terminal sequence and immunoreactivity to Cry10A from Bti.

The 65-kDa protein of 17A is also interesting since, despite having a Bti Cry protein-type N-terminus and being a similar size to both Cry10A and Cry11A it shows very little cross-reactivity to any antisera raised against Bti proteins, again indicating the presence of a novel Cry protein. The N-terminal sequence data obtained provided five possibilities for the ambiguous third residue, T and F (most likely) or I, P and V. Comparison with the other N-termini in Figure 3.10 implies that this residue may be a Proline (P) despite the fact that proline generally gives a very clear sequencing signal.

The mosquitocidal activity of the related strain 84-I, the similarity in polypeptide profile to the highly mosquitocidal strain Bt ssp. *jegathesan* and the identification of Cry type proteins similar to those in the mosquitocidal Bti, all indicated that 17A may also

possess mosquitocidal activity. Pure crystals and spore/crystal mixes were found to be non-toxic to *Aedes aegypti* larvae up to 300 µg/ml of crystal protein. This does not rule out the possibility that 17A may show toxicity to other species of mosquito larvae not tested here.

Four

Chapter 4

Cloning of Novel 65-kDa Putative Toxin

4.1 Introduction

At least 140 holotype Bt insecticidal protein genes have been cloned and characterised (http://www.susx.ac.uk/Home/Neil Crickmore/Bt/index.html), most of which were recognised and identified by the presence of *in vitro* or *in vivo* insecticidal toxicity. These Bt δ-endotoxin genes have been cloned from plasmid vector libraries constructed from Bt total genomic DNA (Held *et al.*, 1982), plasmid DNA (Schnepf & Whiteley, 1981: Ward & Ellar, 1984), or sporulation specific mRNA (Angsuthanasombat *et al.*, 1987). Such libraries can then be expressed in an *E. coli* or *Bacillus* host and screened by immunological cross-reaction (Schnepf & Whiteley, 1981), toxicity (Ward & Ellar, 1984) or by DNA hybridisation with radioactively or non-radioactively labelled probes. Suitable probes can be derived either from other δ-endotoxin gene sequences or based on the protein N-terminal sequence (Granum *et al.*, 1988; Brown & Whiteley, 1992; Van-Nguyen, 1995).

Bt ssp. *fukuokaensis* 17A was isolated from olive groves in Greece (Karamanlidou *et al.*, 1991) and found to be toxic towards the Olive fruit fly *Dacus oleae*. Crystals from 17A were shown to be composed of a complex set of proteins such as is commonly found in diptericidal Bt strains. However, immunological and PCR analysis indicated that the crystal may be composed of a novel set of proteins. N-terminal amino acid sequence analysis revealed that the 17A crystal contained a potentially novel Cry10A-type protein of about 65-kDa. N-terminal analysis of other protein bands present in this strain gave mixed sequence results, but indicated that a second Cry10A-type protein of about 70-kDa may also be present in the crystal.

Cry10A was originally isolated from Bt ssp. *israelensis* (Bti) which is active against mosquito and blackfly larvae. Bti protein inclusions are irregular in appearance (Huber & Lüthy, 1981) and it has generally been accepted that the crystals contain four major δ-endotoxin proteins with molecular weights of 125-kDa (Cry4A), 130-kDa (Cry4B), 68-kDa (Cry11A) and 27-kDa (Cyt1A) (Donovan *et al.*, 1988; Federici *et al.*, 1990; Höfte & Whiteley, 1989). A further δ-endotoxin of 72-kDa called Cry10A was originally thought to be a minor component of the parasporal body (Thorne *et al.*, 1986; Garduno *et al.*, 1988), but has since been identified as a fifth major crystal component (Purcell, 1997). Cry4A, Cry4B and Cry11A δ-endotoxins are also found in the highly mosquitocidal strain Bt ssp. *morrisoni* PG14 along with Cyt1A and a 144-kDa protein (Earp *et al.*, 1987; Ward *et al.*, 1984).

More recently, two more δ-endotoxin genes have been identified and cloned from the Bt mosquitocidal strains *jegathesan* (Delécluse *et al.*, 1995) and *medellin* (Restrepo *et al.*, 1997). Bt ssp. *jegathesan* was found to harbour a gene, designated *cry11B*, encoding an 81-kDa crystal protein (Delécluse *et al.*, 1995). The Cry11B protein sequence contains large regions of similarity to that of Cry11A and is immunologically related to both Cry11A and Cry4A δ-endotoxins. Purified Cry11B inclusions were highly toxic to mosquito larvae of the species *Aedes aegypti, Culex pipiens* and *Anopheles stephensi* and showed similar activity to that of the native Bt ssp. *jegathesan* crystal, which contains at least seven polypeptides (Ragni *et al.*, 1996). The gene encoding a 94-kDa protein from Bt ssp. *medellin* has also been cloned (Restrepo *et al.*, 1997). Purified inclusions from the recombinant strain were found to be toxic to *Aedes aegypti* and *Culex quinquefasciatus*, although the toxicity was not as high as that produced by the native crystal which contains additional proteins of 68 and 28-kDa (Restrepo *et al.*, 1997).

The identification of novel diptericidal proteins may be valuable, both in the development of pesticides with novel specificities, and also in delaying or even preventing resistance development to currently used formulations. It has been found that combining various toxins with similar specificities in one recombinant cell may prevent or delay the onset of resistance (Georghiou *et al.*, 1993, 1997). On a less practical note, the similarities between Bt diptericidal proteins from the different strains described above provides an interesting study of the evolutionary origins of both the δ-endotoxins and their host strains. A recent study carried out by Bravo (1997) on the phylogenetic relationships of the Cry family of proteins made several observations on the evolutionary origins of Bt biocidal activity and the roles and origins of the different toxin domains, as well as commenting on possible additional biocidal activities for some proteins. It was also suggested that knowledge of the evolutionary relationships of the different toxin domains may set the basis for a more rational and directed strategy to create novel chimaeric toxins that have different specificities (Bravo, 1997). The cloning and sequence analysis of additional diptericidal proteins will therefore enable a fuller study of their phylogeny that may also yield information useful for future genetic engineering projects.

This Chapter describes the cloning of plasmid DNA fragments believed to harbour the gene for a novel Cry10A-type 65-kDa protein from Bt ssp. *fukuokaensis* 17A. Size enriched plasmid DNA libraries were constructed and screened using DIG (digoxygenin)-labelled oligonucleotides based upon the N-terminal sequence of the 65-kDa protein. DIG is a comprehensive, convenient and effective system for the labelling and detection of DNA, RNA and oligonucleotides. The use of a non-radioactive probe confers several advantages; namely that the technology is safe, the probes can be stored for at least a year and the hybridisation solutions can be reused several times.

In principle, cloning into plasmid vectors is straightforward. The plasmid DNA is cleaved with a suitable restriction enzyme and joined *in vitro* to a fragment of foreign DNA. The resulting constructs are then used to transform bacteria. In practice however, the plasmid vector must be carefully chosen to minimise the effort required to identify and characterise recombinants. The selection of bacteria containing plasmid DNA is easily and routinely achieved by using a plasmid vector containing an antibiotic-resistance gene. Transformed cells are then plated onto media containing that antibiotic. It is more difficult however to distinguish between plasmids that contain inserted foreign DNA and vector molecules that have recircularised without any insertion. In the plasmid vector pGEM-3Zf(+) (the Promega vector used in this cloning procedure) this selection is facilitated by the presence of a sequence coding for the lac α-peptide interrupted by a multiple cloning site. Non-recombinant plasmids produce a functional α-peptide which, by complementation of the host cell *lacZ*ΔM15 gene product, leads to production of functional β-galactosidase. Bacterial colonies harbouring the *lacZ*ΔM15 gene on an F' and also containing a pGEM-3Zf(+) vector are blue when plated on indicator media containing IPTG and X-Gal. However, when the lac α-peptide is disrupted by cloning into the pGEM-3Zf(+) multiple cloning region, complementation does not occur and no β-galactosidase activity is produced. Bacterial colonies harbouring recombinant pGEM-3Zf(+) vector constructs are therefore white.

4.2 Design and Synthesis of the Oligonucleotide Probes

In an attempt to clone the gene encoding the 65-kDa protein from Bt ssp. *fukuokaensis* 17A, synthetic oligonucleotide probes were designed. These oligonucleotides were based on the N-terminal amino acid sequence described in the previous Chapter (Figure 3.10) after consideration of Bt δ-endotoxin codon usage (Figure 4.1). It was hoped that the similarity of the 17A 65-kDa protein N-terminal sequence to N-terminal sequences obtained from the ambiguous 70-kDa protein band might enable the detection of two or more related proteins from this strain using a single probe designed to the 65-kDa sequence.

Figure 4.2 shows the design of two 17A oligonucleotides with respect to the N-terminal sequence of the 65-kDa protein. The first oligonucletide (17A.1) is non-degenerate and based on the codon usage tables whereas the second oligonucleotide (17A.2) contains three degeneracies to optimise the similarity to the *cryIVA* N-terminal sequence.

The single-stranded oligonucleotides were labelled by incorporation of digoxygenin (DIG) using a terminal transferase (Section 2.2.9). This was achieved by labelling the

A GCX (75)	C TGT (70)
D* GAT (80)	E* GAA (70)
F* TTT (80)	G* GGA (50), GGT (25)
H CAT (80)	I* ATT (50)
K* AAA (75)	L TTA (50)
N* AAT (80)	P CCT (30), CCA (50)
Q CAA (90)	R* AGA (40), CGT (20), AGG (20)
S TCN (75), TCX (50), AGT (25)	T* ACA (45), ACT (30)
V* GTX (75)	Y* TAT (80)

Figure 4.1 Codon usage by Bt δ-endotoxins (Van-Nguyen, 1995).
Codon usage was analysed using *cyt1A, cry1Aa, cry1Ba, cry1Ca, cry2Aa, cry2Ab, cry3Ba, cry4A* and *cry4B*. All were found to have very similar codon usage, except the *cry2* δ-endotoxins which differed substantially from the others in the residues that are shown with an asterisk alongside. Where there is clearly not a dominant codon, the alternatives are shown. The codons for each residue are followed by the average percentage occurrence in brackets. The letters X and N in the codon sequences indicate "A or T" and "any base", respectively.

Figure 4.2a Oligonucleotide 17A.1

```
5'                                                    3'
 M    N    P    Y    Q    K    K    N    E
ATG AAT CCA TAT CAA AAA AAA AAT GAA
```

Figure 4.2b Oligonucleotide 17A.2

```
5'                                                    3'
 M    N    P    Y    Q    K    K    N    E
ATG AAT CCA TAT CAA AAA AAA AAT GAA
         T              T    G
```

Figure 4.2 N-terminal sequence of the 65-kDa protein from Bt ssp. *fukuokaensis* 17A and the corresponding 24-base oligonucleotides (17A.1 and 17A.2).
Selected to screen for transformants and designed by deduction from the codon usage table (Figure 4.1). The oligonucleotide 17A.2 is degenerate at three positions (9, 18 and 21) lowering the predicted T_m by 3°C.

oligonucleotides at the 3' end by incorporation of dATP and DIG-11-dUTP to form a 3' extension of 40-50 bases with several DIG molecules incorporated. Each DIG-labelled probe can then be detected after hybridisation by a high-affinity anti-digoxygenin-antibody Fab fragment conjugated to alkaline phosphatase (AP). The alkaline phosphatase dephosphorylates the chemiluminescent substrate Disodium 3-(4-methoxyspiro{1,2-dioxetane-3,2'-(5'-chloro)tricyclo[3.3.1^{3,7}]decan}-4-yl)phenyl phosphate, (CSPD), which, upon decomposition *via* an unstable intermediate, emits light. This is then detected on X-ray film (Section 2.2.12).

Full details of the methodology for labelling the two probes with DIG-11-dUTP are described in Section 2.2.9. The success of each labelling reaction was confirmed by direct detection of the tailed oligonucleotide probes in a dot blot experiment (Section 2.2.9) (Figure 4.3). Comparison with a pre-labelled control also enabled the yield of DIG-labelled oligonucleotide to be estimated (Figure 4.3). The concentrations of DIG-labelled 17A.1 and 17A.2 were both found to be around 1 pmol/µl.

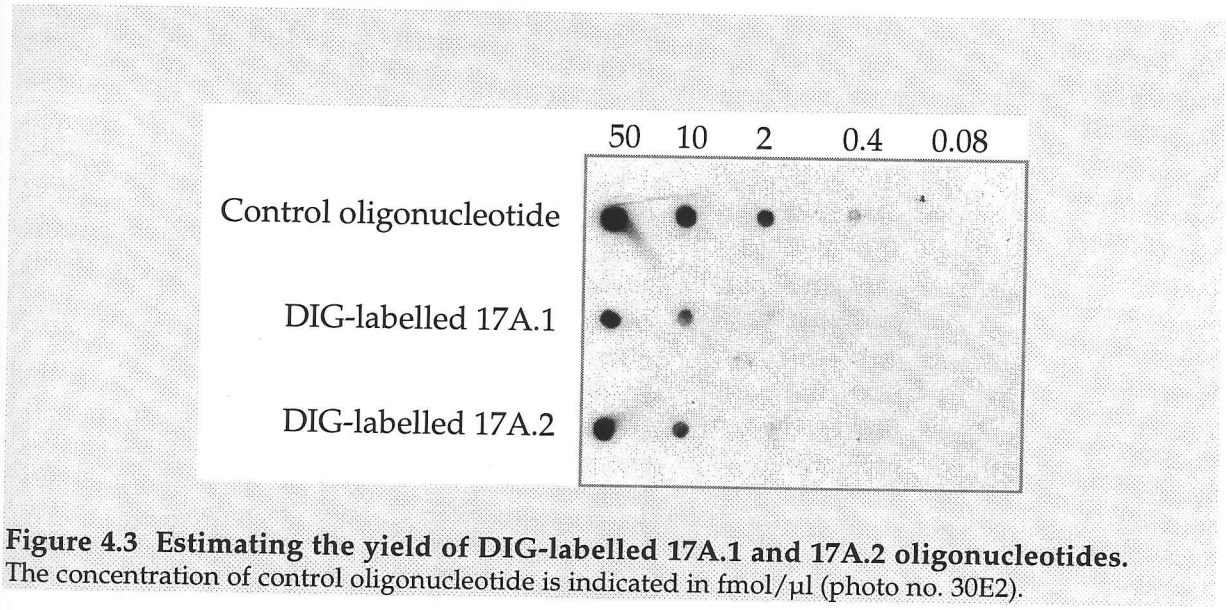


Figure 4.3 Estimating the yield of DIG-labelled 17A.1 and 17A.2 oligonucleotides.
The concentration of control oligonucleotide is indicated in fmol/µl (photo no. 30E2).

4.3 Bt ssp. *fukuokaensis* 17A DNA Hybridisation Studies and Plasmid Library Construction

Bt δ-endotoxins are generally found on plasmids (Gonzalez *et al.*, 1981; Carlton & Gonzalez, 1985). This extrachromosomal location was first indicated by loss of crystal production in strains that had also lost a plasmid or plasmids (Gonzalez *et al.*, 1981) and later confirmed by Kronstad *et al.* (1983) when the δ-endotoxin genes of 14 different Bt subspecies were located on plasmid DNA. It would be possible to clone the 65-kDa protein gene from a 17A total DNA library; however extensive screening would be needed to

eliminate clones containing inserts of chromosomal DNA. Isolating 17A plasmid DNA thereby provides a means to facilitate the cloning of this gene.

Dunn (1996) had reported difficulty in preparing total plasmid DNA from Bt ssp. *fukuokaensis* 84-I using the alkaline lysis method that had been successfully used for other Bt strains (Birnboim & Doly, 1979; Sambrook *et al.*, 1989; Van-Nguyen, 1995). Plasmid preparations from 17A using this method did prove to be of low yield, as determined by DNA Gene Quant (typically 100 µg per litre of culture) and inconsistent quality as determined by agarose gel electrophoresis. It was observed that the key step in obtaining plasmid DNA of good restriction enzyme digest quality was resuspension of the large, sticky isopropanol pellet. Any attempt to resuspend the pellet manually with even a large bore pipette caused a decrease in quality of the resulting DNA; this was presumably due to shearing of the large plasmid or plasmids leaving them more suceptible to degradation.

The method found to give the best results was to wash the pellet from the side of the centrifuge tube with 10x TE into a 15 ml capped tube, tape the tube to a gentle rocker and leave the pellet to soak and soften in the 10x TE for at least two hours. The pellet was then gently broken up with a small spatula and the tube left on the rocker at room temperature for several hours and preferably overnight until the pellet had completely resuspended. The plasmid DNA was then separated from contaminating chromosomal DNA and RNA on a continuous caesium chloride gradient (Section 2.2.2b) and propan-2-ol precipitated. Even with the above precautions the isolated plasmid DNA was still contaminated with chromosomal DNA that appears as smearing in restriction digests (see Figure 4.5a). Attempts to purify further the plasmid DNA by subjecting it to a second caesium chloride gradient proved unsuccessful as no band was recovered; this was probably due to the initial yield of plasmid DNA being too low to survive the losses that accompany each purification step.

Despite the difficulty in obtaining plasmid DNA free of chromosomal contamination such preparations are certainly plasmid enriched. Plasmid libraries contructed from these DNA preparations should therefore contain a higher ratio of clones containing the desired insert fragment than libraries constructed from total cell DNA thereby reducing the size of the plasmid library to be screened. It was therefore decided to use these preparations of 17A plasmid DNA from which to clone the putative Cry toxin.

Caesium chloride-gradient purified Bt ssp. *fukuokaensis* 17A DNA was digested with several restriction enzymes (5 µg per reaction), (Section 2.2.3) electrophoresed on an 0.8% agarose gel and Southern blot analysed with the two digoxygenin-labelled oligonucleotide probes (17A.1 and 17A.2, degenerate and non-degenerate respectively) (Sections 2.2.10 to 2.2.13). The DNA size markers used were the *Hin*dIII digested and DIG-tailed lambda phage DNA supplied by Boehringer. Separate blots were hybridised with each probe. Initially the stringent wash step was carried out at 0.1x SSC (15 mM NaCl; 1.5 mM sodium

citrate, pH7) as stated in the Boehringer Mannheim protocol and the predicted hybridisation temperature calculated according to the equation in Figure 4.4. The melting temperature (T_m) for each probe was calculated to be 56°C for 17A.1 and 53°C for 17A.2, where the three degeneracies reduced the T_m by 3°C.

$$T_m = 81.5 + 16.6 (\log_{10}[Na^+]) + 0.41 (\%G+C) - 600/N$$

T_m = melting temperature
$[Na^+]$ = concentration in final wash.
N = length of oligonucleotide.

Figure 4.4 Equation used to calculate the predicted T_m values for the DIG-labelled oligonucleotide probes 17A.1 and 17A.2.

The hybridisation and final stringent wash steps were therefore carried out at 47°C (5-10°C below the T_m). When no hybridisation was seen for either probe the hybridisations were carried out at successively reduced temperatures but no positive signals were seen even at room temperature. The stringency was then lowered further by increasing the SSC concentration to 1x and the hybridisations carried out at 30°C. Under these conditions several sized bands of DNA then hybridised with each probe. Further experiments with the temperature and stringency of the final wash step finally identified conditions of 0.5x SSC and 48°C as giving the clearest pattern of hybridisation (Figure 4.5a). This increase in stringency gave two main hybridising bands and one or two fainter bands for each digest. Since the degenerate and non-degenerate probes gave the same result the non-degenerate probe, 17A.1, was used in all subsequent experiments.

Two hybridising bands were deemed to be acceptable since the probe was designed in the hope that it could be used to identify more than one 17A toxin gene. The probe was found to hybridise most strongly to restriction fragments of the approximate sizes shown in Figure 4.5b.

The two *Hin*dIII fragments were selected for an initial attempt at cloning the novel gene or genes. A concurrent attempt was also made to clone the two *Eco*RI fragments but this was unsuccessful and is not discussed in this report. The 4.4 kb *Hin*dIII fragment is large enough to contain the whole of the approximately 2.0 kb of DNA necessary to encode a 65-kDa protein. The 0.7 kb *Hin*dIII fragment is obviously to small to contain the whole gene but its small size is convenient for rapid sequencing and identification of the hybridising region of DNA.

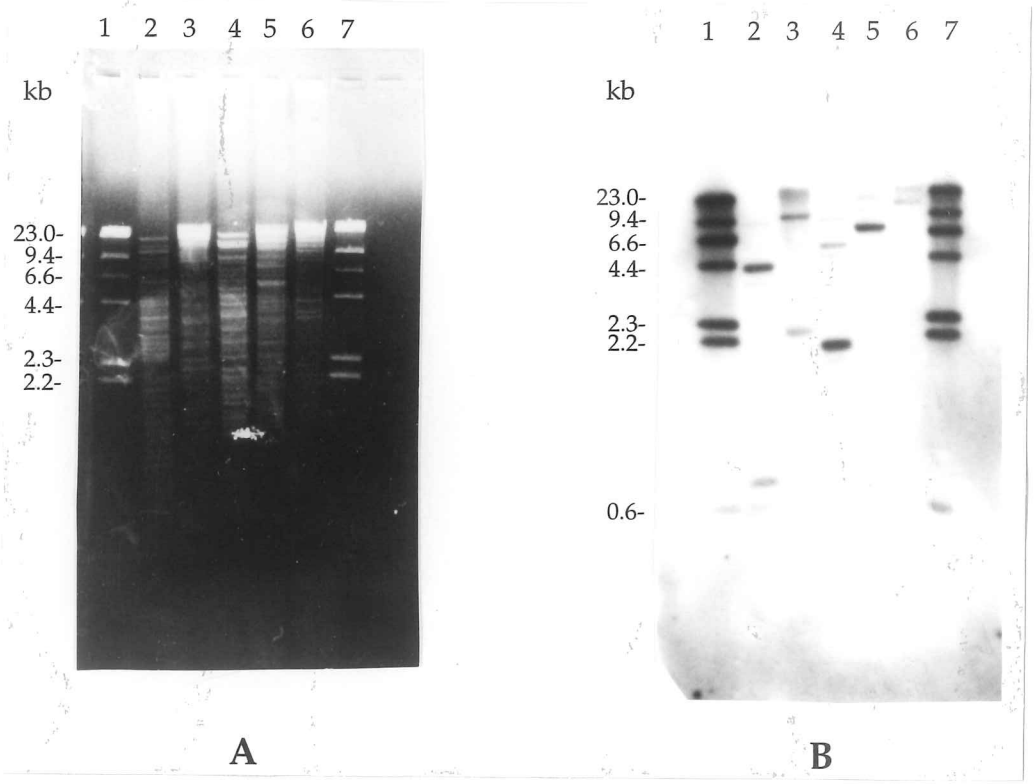


Figure 4.5a Southern blot analysis of Bt ssp. *fukuokaensis* 17A plasmid DNA (photo no. 980D1).
Panel A: 0.8% agarose gel stained with ethidium bromide showing caesium chloride purified Bt ssp. *fukuokaensis* DNA digested with *Hin*dIII (lane 2), *Pst*I (lane 3), *Eco*RI (lane 4), *Sca*I (lane 5) and *Sph*I (lane 6) (5µg/lane). Digoxygenin-labelled λ-*Hin*dIII markers were run in lanes 1 and 7.
Panel B: Southern blot and hybridisation of Panel A with digoxygenin-labelled 17A.1 (final stringent wash 0.5x SSC/48°C).

Restriction enzyme	Large fragment	Small fragment
*Hin*dIII	4.4 kb	0.7 kb
*Pst*I	10.0 kb	2.1 kb
*Eco*RI	6.0 kb	2.0 kb
*Sca*I	21.0 kb	7.0 kb
*Sph*I	23.0 kb	15.0 kb

Figure 4.5b Approximate sizes of the restriction fragments of 17A plasmid DNA that bind the 17A.1 olignonucleotide probe.

In order to clone the *Hin*dIII fragments, caesium chloride-gradient purified Bt ssp. *fukuokaensis* 17A DNA (100 µg) was digested with *Hin*dIII in a total volume of 200 µl for eight hours. The digested DNA was then electrophoresed on a 0.5% agarose gel in 1x TAE at 40 volts for eighteen hours. Four slices of gel were excised from the digest in the 3-5 kb region and four from the 0.6-0.8 kb region, the DNA from each slice was repurified using a Geneclean II kit (Section 2.2.5) and resuspended in a volume of 20 µl. Initial cloning attempts involved running up to half of each sample on a 0.5% agarose gel to check the recovery and purity of each fraction. This gel was then subjected to Southern blot analysis using the DIG-labelled 17A.1 oligonucleotide probe to identify the fraction containing the highest concentration of hybridising DNA. The remaining portion of that hybridising fraction was then used in a subsequent ligation reaction with the selected vector. However, it was found that recovery of the DNA restriction fragments was low and the fraction remaining after half had been used in Southern blot analysis contained insufficient DNA for successful ligations. Eventually eight individual ligations were attempted using the total volume of DNA recovered from each gel slice. Ligations carried out in this manner were much more successful, as is described below, although screening of a considerable number of transformants was then necessary.

The plasmid pGEM-3Zf(+) (Promega) (Figure 4.6) was selected as the most suitable vector in which to clone the above fragments as it has an extensive restriction endonuclease recognition polylinker (useful for endonuclease mapping and subcloning) and the *lacZ* gene (allowing blue/white selection for recombinant colonies) as described in the introduction to this Chapter.

In order to clone the *Hin*dIII fragments, total recovery of the insert preparations were ligated to about 15 ng of *Hin*dIII linearised, dephosphorylated pGEM-3Zf(+) (approximately a 1:3 molar ratio of vector to insert) at 16°C in a total volume of 20 µl for sixteen hours. Removal of the 5'-phosphate groups with alkaline phosphatase (Section 2.2.6) was used to suppress self-ligation and circularisation of the plasmid DNA. During ligation *in vitro*, bacteriophage T4 DNA ligase catalyses the formation of a phosphodiester bond between adjacent nucleotides only if one nucleotide contains a 5'-phosphate group and the other contains a 3'-hydroxyl group. Recircularisation of plasmid DNA is therefore minimised by removing the 5' phosphates from both termini of the linear DNA with calf intestinal alkaline phosphatase (CIP) (Sambrook *et al.*, 1989). When the plasmid DNA has been dephosphorylated only two new phosphodiester bonds can be formed to give an open circular molecule containing two nicks. Because circular DNA (even nicked circular DNA) transforms much more efficiently than linear plasmid DNA, most of the transformants will contain recombinant plasmids. The resulting hybrid molecules carry two single-stranded nicks that are repaired after the hybrids have been introduced into competent bacteria.

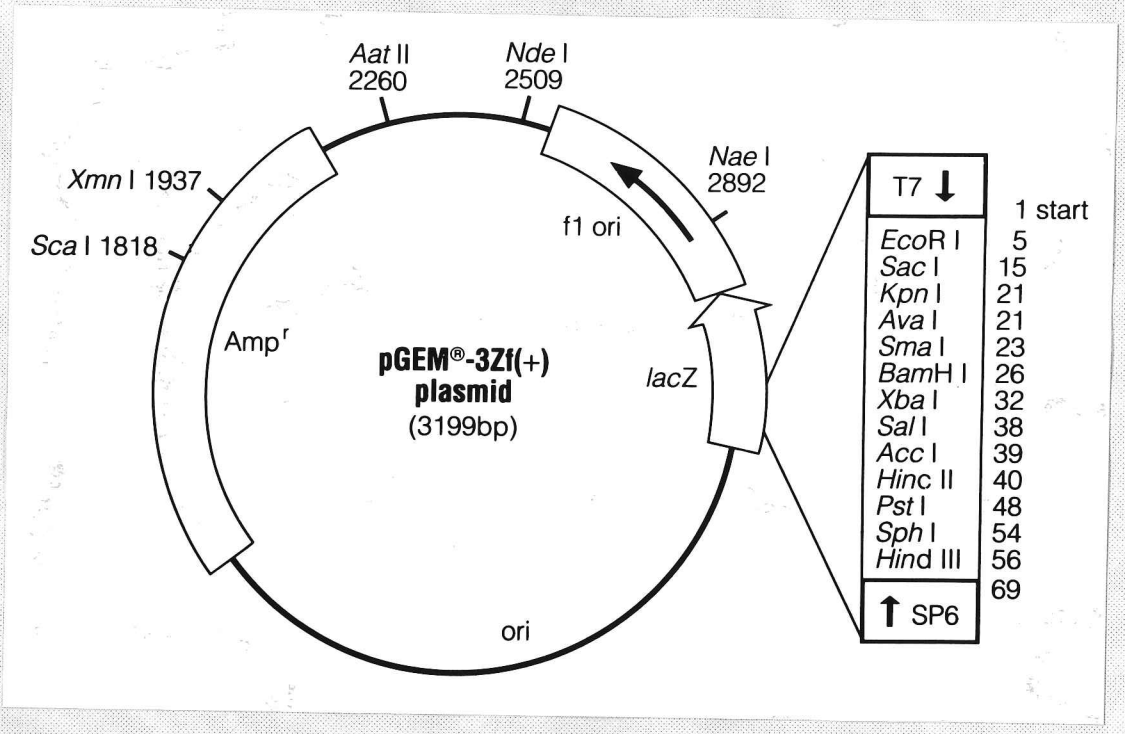


Figure 4.6 pGEM-3Zf(+) vector circle map.
Taken from the Promega Protocols and Applications Guide (second edition) (photo no. 29E6).
The T7 and Sp6 RNA polymerase initiation sites are at positions 1 and 69, respectively, on either side of the multiple cloning sites polylinker. The ampicillin-resistance gene (β-lactamase) is represented as Amp^r and the f1 origin and its orientation for the production of single-stranded DNA are also indicated.

Ligation of the segment of foreign DNA to the linearised plasmid vector involves the formation of new bonds between phosphate residues located at the 5' termini of double-stranded and adjacent 3'-hydroxyl moieties. During the ligation reaction, the insert DNA has the capacity to oligomerize (this would also occur with the vector plasmid if the 5' phosphates had not been removed) reducing the concentration of insert DNA available to form the desired recombinant molecule with plasmid DNA. Oligomerisation of the insert DNA was reduced by heating to 50°C for ten minutes to break down any oligomers then rapidly cooling on ice directly before the other components of the ligation reaction were added.

Following ligation (Section 2.2.8) the reaction mixtures were incubated at 65°C for ten minutes to inactivate the T4 DNA ligase which would otherwise interfere with the electroporation efficiency. Electroporation is a method routinely used to transform *E. coli* (Dower *et al.*, 1988) and other bacteria (Chassy & Flickinger, 1987; Bone & Ellar, 1989). In this case 1 µl of each reaction was electroporated (Section 2.2.16a) into 50 µl of *E. coli* JM109 in a 4 mm cuvette at 1.5 kV and 1000 Ω. This procedure was carried out at low

temperatures (0-4°C), as the efficiency of transformation drops as much as 100-fold when carried out at room temperature.

More than 500 ampicillin-resistant colonies were produced for each of the eight ligation reactions of which approximately 90% were white upon LacZα colour selection. Sixteen white colonies were picked from each of the eight plates and used to inoculate 5 ml of LB containing 100 µg/ml ampicillin (four colonies per culture, thirty-two cultures in total). The resulting cultures were used in small scale plasmid preparations (Section 2.2.1a) and the DNA subjected to Southern blot analysis with the DIG-labelled 17A.1 probe. Of these thirty-two bulk preparations, three gave positive hybridisation signals, two corresponding to the 4.4 kb fragment and one to the 0.7 kb fragment. Small scale plasmid preparations were then made from the twelve individual colonies that made up the three positively hybridising bulk preparations and the DNA subjected to a second round of hybridisation with the 17A.1 probe. Three positive hybridisation signals were detected (one for each group of four preparations). One clone containing the 4.4 kb fragment (pAM4.4) and the clone containing the 0.7 kb fragment (pAM0.7) were selected for further study.

The pGEM-3Zf(+) plasmid contains priming sites for the universal primers T7 and Sp6 on either side of the polylinker (see Figure 4.6) which can be used to obtain the sequence for any DNA cloned into one of the polylinker cloning sites. Using an automated sequencing reaction and good quality DNA it is therefore possible to generate up to 0.6 kb of reliable sequence from each end of a cloned insert. In order to obtain complete sequence in both directions for insert fragments larger than 0.6 kb a more thorough sequencing strategy is required and this is discussed in Chapter 5. However, if part of the gene of interest lies within the first 0.6 kb of either end of the insert this technique can provide a rapid means of confirming that the clone does indeed contain the desired gene as well as indicating its orientation and location within the insert.

pAM4.4 and pAM0.7 DNA was prepared using the Perfect prep™ plasmid DNA kit (2.2.1b) which produces DNA of a suitable quality and concentration for automated sequencing and 1.2 kb of unambiguous sequence was obtained for each clone (0.6 kb from either end of the insert). In the case of pAM0.7 the sequences were of course overlapping so sequence was obtained for the whole of the insert fragment. The two overlapping sequences were matched by the "bestfit" program from the GCG package of programs, which allows comparison of sequence data from different strands, and compiled to produce the 715 bp of unambiguous sequence shown in Figure 4.7.

A six-phase translation of this sequence identified the initial 37 amino acids of an open reading frame beginning with the amino acid sequence MNSYQNKNE, subsequently referred to as *orfA* (Figure 4.7). Comparison of this sequence to the N-terminal amino acid sequence derived for the 65-kDa protein from 17A showed identity to seven of the nine

amino acids. Analysis of the corresponding DNA sequence identified a region differing in only three nucleotides (Figure 4.8) from that of the 17A.1 probe.

The DNA and amino acid sequence of the clone pAM0.7 was analysed using the BLAST database search from the GCG package of programs. Significant levels of similarity and identity were found to several δ-endotoxin genes, most notably Cry4A from Bt ssp. *israelensis* (about 35% identity) indicating that part of at least one novel δ-endotoxin gene had been cloned from Bt ssp. *fukuokaensis* 17A. A restriction map of clone pAM0.7 is shown in Figure 4.9.

```
T7---->
AAG CTT TAA GTC GGA TTT TTA TTT ATA ATT TTT TAT AAA GTT TCC CTA AAT CTT TCT TTA 60

TCT GAG ATT TCA GTT GAT GAT TTT CGT CTT CTA GTG TCT TAA CTT TCT GTC TTA GTA GCT 120

TCA TTT TAT CTT TTA AAG TTT CTA TAA TTA CAG TTT TCT GTA CAC AGG TAA AAT TCC CAT 180

AAT TCG GTT TTT CCT AGA AAA AAT AAA ATG AAT CCA ATA CAG GTT CGG ATT TAT TTT CTA 240

CAA TTT CTT TCA ATG ACG AAT AAT TAA ATA TCT ACA ACA CTC TAT GTA GTA TCC CTC TCT 300

TTT TTA GAG GGG GGG AAA AAA CAT CCA AGG GTG AAT TTC GTT CAT ATA AAA GTG AAT GAC 360

TCT TTT CGC ACC CTT AAA AAC AAC AAA GAA AAA ATC GTT CTA TAG AAA TCT GAA ACT TCA 420

AAA AAT TAT ATG CAA TAC ATA AAG AAA AGG TTT AAA AAA TCA ATA TTT TTA CCA AAA ATA 480

ATG GGT TTA TTT GTA GAA ACA TTG TCA CAG GAA TAC ATT GGG GCA CTA CGA ATA TAT AAC 540

AAG ACA CCT AGC ATA TAC TGT TTG GGT ATC TAA AAA TAA GGA CCA CAT AAG GGA GTG AAA 600

    M   N   S   Y   Q   N   K   N   E   Y   E   I   L   D   A   S   R   N   N
AAT ATG AAT TCA TAC CAA AAT AAA AAT GAA TAT GAA ATA TTG GAT GCT TCA CGA AAC AAC 660

S   N   M   S   N   R   Y   P   R   Y   P   L   A   N   D   P   D   A
TCT AAT ATG TCT AAC CGT TAT CCA AGG TAT CCA CTA GCA AAT GAT CCA CAA GCT T

                                                                    <----Sp6
```

Figure 4.7 Nucleotide sequence of the insert DNA from clone pAM0.7.
The open reading frame *orfA*, described in the text is shown using the one letter amino acid code. The *Hin*dIII restriction enzyme sites used to construct the clone are underlined and the direction of the sequence obtained from the T7 and Sp6 primers is indicated.

```
17A.1          M   N   P   Y   Q   K   K   N   E
               ATG AAT CCA TAT CAA AAA AAA AAT GAA

OrfA           M   N   S   Y   Q   N   K   N   E
               ATG AAT TCA TAC CAA AAT AAA AAT GAA
```

Figure 4.8 Comparison of the amino acid and nucleotide sequences of the oligonucleotide probe 17A.1 and the initial nine codons of OrfA from pAM0.7.
The three nucleotide discrepancies between the two sequences are in bold.

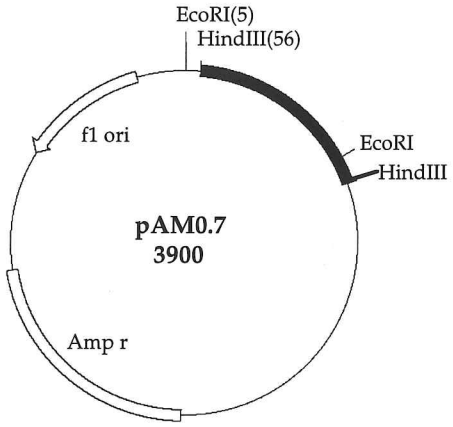


Figure 4.9 Restriction map of plasmid construct pAM0.7.
The 0.7 kb fragment (black box) is cloned into the *Hin*dIII site of the vector pGEM-3Zf(+). The ampicillin resistance gene is represented as Amp r and the f1 origin is indicated. Positions 5 and 56 mark the limits of the pGEM-3Zf(+) polylinker. The sequence shown in Figure 4.7 runs clockwise in this map.

The 1.2 kb of sequence generated from clone pAM4.4 was also subjected to a BLAST database search. No significant homology was found to any previously identified δ-endotoxin genes. Neither was any sequence found that would be likely to bind the 17A.1 probe. However, 3.2 kb of DNA still remained unsequenced, sufficient to contain the whole of a second δ-endotoxin gene of about 60 to 70-kDa. One positive result generated by the sequence data was the identification of a low G+C ratio compatible with that of Bt DNA. This helped confirm that the fragment of DNA contained in the construct was indeed from 17A and not from a contaminating organism as sometimes occurs.

In order to locate more accurately the hybridising region and identify suitable fragments for subsequent subcloning experiments, restriction mapping was employed on construct pAM4.4 using the enzymes with sites present in the pGEM-3Zf(+) polylinker (Figure 4.10). This was followed by Southern blotting and hybridisation using the 17A.1 probe. A restriction map of pAM4.4 is shown in Figure 4.11; the hybridising region is believed to lie between the *Hin*cII and the *Pst*I sites. Digestion with *Pst*I resulted in the generation of two fragments of 5.0 and 2.6 kb, the smaller of which hybridised the 17A.1 probe. This 2.6 kb *Pst*I fragment was selected as suitable for subcloning to facilitate the generation of further sequence information, and this is described in Chapter 5.

Cloning of the 0.7 kb *Hin*dIII fragment from 17A plasmid DNA had positively identified the existence of at least one δ-endotoxin gene in Bt ssp. *fukuokaensis* 17A (*orfA*). However, this clone only contained the first 37 amino acids of this toxin gene. The next stage was to attempt to clone the whole of *orfA* on a larger fragment of DNA, and for this purpose the two large 15 kb and 23 kb *SphI* fragments (see Figure 4.5) were selected.

Digestion of 17A DNA with *Sph*I, preparation of the insert and vector DNA and subsequent ligations were carried out as previously described for the cloning of the *Hin*dIII fragments. This time there were only about 50 colonies per transformation, of which about 40% were white. The lower ligation efficiency was probably due to the large size of the insert DNA. Bulk preparations of several white colonies were carried out as previously

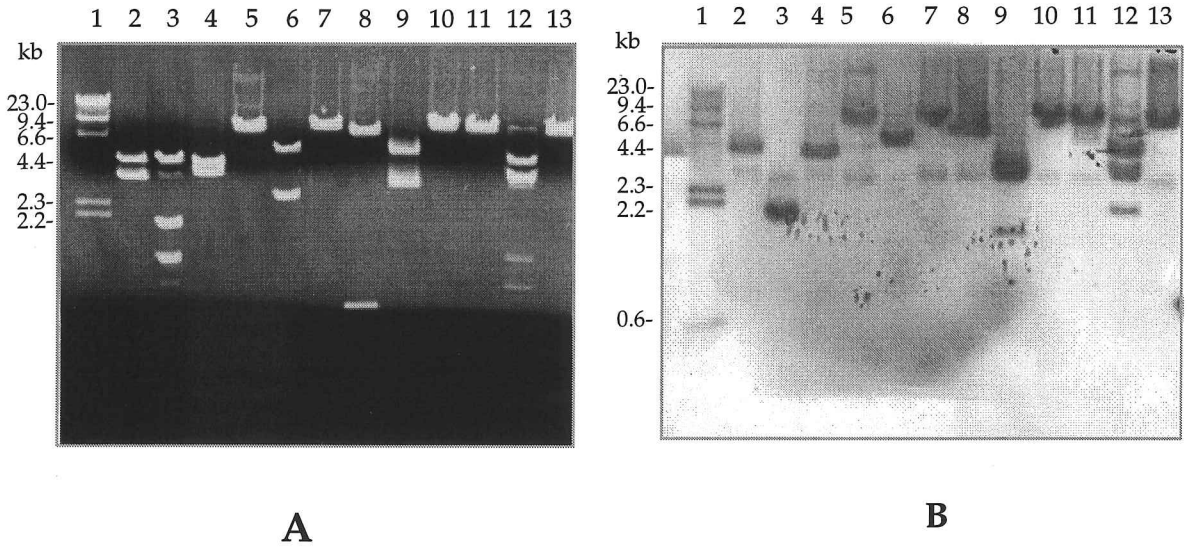


Figure 4.10 Restriction endonuclease analysis of clone pAM4.4 (photo no. 30E1).
Panel A: 0.8% agarose gel stained with ethidium bromide showing construct pAM4.4 digested with the following enzymes. The size marker in lane 1 is *Hin*dIII-digested DIG-labelled lambda DNA.

Lane 1	Size marker	Lane 8	*Hin*cII
Lane 2	*Hin*dIII	Lane 9	*Pst*I
Lane 3	*Eco*RI	Lane 10	*Sph*I
Lane 4	*Acc*I	Lane 11	*Kpn*I
Lane 5	*Xba*I	Lane 12	*Bam*HI
Lane 6	*Sca*I	Lane 13	*Sal*I
Lane 7	*Sma*I		

Panel B: Southern blot and hybridisation of Panel A with DIG-labelled 17A.1.

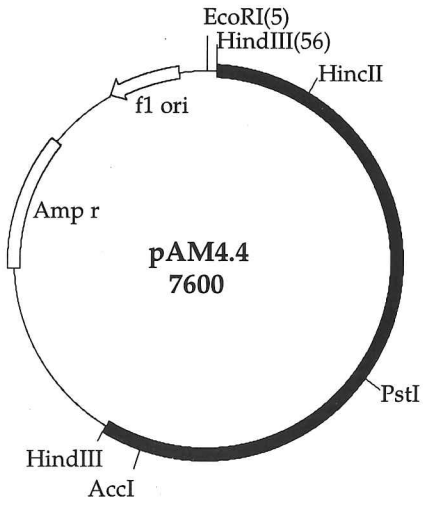


Figure 4.11 Restriction map of plasmid construct pAM4.4.
The 4.4 kb fragment (black box) from Bt ssp. *fukuokaensis* 17A plasmid DNA was cloned into the *Hin*dIII site of the vector pGEM-3Zf(+). The ampicillin-resistance gene is represented as Amp r and the f1 origin is also indicated. Positions 5 and 56 mark the limits of the pGEM-3Zf(+) polylinker. The sequence shown in Figure 4.7 runs clockwise in this map.

described and the DNA subjected to Southern blot hybridisation. One positively hybridising preparation was obtained containing a 15 kb insert. A second round of hybridisation on DNA from the individual colonies identified two positively hybridising preparations. One of these, pAM15, was selected for further investigation.

Sequence from each end of the clone was generated using the T7 and Sp6 universal primers as previously described. A six-phase translation of sequence obtained from the T7 primer identified an open reading frame running the entire length of the sequence on the complementary strand; this open reading frame is subsequently referred to as *orfB* and is shown in Figure 4.12. BLAST searches carried out on the DNA and amino acid sequences of *orfB* identified significant similarity to sequence from the Cry4A δ-endotoxin of Bt ssp. *israelensis* (discussed in Chapter 5). The similarity was found to the C-terminal region of a *cry4A* gene running in an orientation that suggested that the N-terminal region of the gene lay within the body of the clone. This is consistent with the N-terminal region acting as a recognition site for the 17A.1 probe. Sequence data obtained from the Sp6 primer did not show any similarity to any previously cloned δ-endotoxin genes.

The discovery of δ-endotoxin sequence on the edge of construct pAM15 showed that the whole of the gene could not be contained within the cloned *Sph*I fragment. However, it was decided to proceed with sequence analysis of that part of the gene which did lie within the clone in order to identify suitable restriction sites that could be used to clone the complete gene. It was also necessary to identify and sequence the N-terminus so that this sequence could be compared with that obtained for *orfA* from pAM0.7 to determine if they were parts of the same gene.

Due to the large size of the cloned fragment, full restriction analysis was carried out in order to characterise the construct and identify fragments suitable for generating smaller subclones to facilitate sequence analysis (Figure 4.13). *Hin*dIII digestion resulted in the production of eight restriction fragments of sizes ranging from 0.6 kb to about 4.6 kb including a 0.7 kb fragment. *Eco*RI digestion resulted in the production of four restriction fragments of about 5.3 kb, 4.8 kb, 2.5 kb and 1.9 kb. Subsequent blotting and hybridisation revealed that the 0.7 kb *Hin*dIII fragment and the 2.5 kb *Eco*RI fragment contained the hybridising sequence.

The identification of a 0.7 kb *Hin*dIII fragment provided further evidence that the gene contained in construct pAM15 is the same as that in pAM0.7. This fragment was therefore selected for subsequent subcloning and sequence analysis so that this could be confirmed. Reference to the pAM0.7 restriction map in Figure 4.10 and sequence data in Figure 4.7 suggests that the 2.5 kb *Eco*RI fragment would start within the first codon of *orfA*. If this is the case then the whole of a gene corresponding to a 65-kDa protein (about 2 kb) could lie within this fragment. This deduction is not consistent with the pAM15 sequence

T7--->

```
GCA TGC ATA CCT ATC AGA AGC ATA TCC ATC TGT TAC TGG CGT AAG ATA ACC TTG GCT GAC 60
CGT ACG TAT GGA TAG TCT TCG TAT AGG TAG ACA ATG ACC GCA TTC TAT TGG AAC CGA CTG
C   A   Y   R   D   S   A   Y   G   D   T   V   P   T   L   Y   G   Q   S   V

ATA CTG GGA CGA TGA TTC ACA GCG ATA ATC TCC ACA TGA AGG ATT AGG CTG CAT ATA GGA 120
TAT GAC CCT GCT ACT AAG TGT CGC TAT TAG AGG TGT ACT TCC TAA TCC GAC GTA TAT CCT
Y   Q   S   S   S   E   C   R   Y   D   G   C   S   P   N   P   Q   M   Y   S

CAA ATC ATT TGG AAC ATT CAT AAT GGC GTC AAT TTC TTT CCC GTA ACG TGT TAC CAC TAA 180
GTT TAG TAA ACC TTG TAA GTA TTA CCG CAG TTA AAG AAA GGG CAT TGC ACA ATG GTG ATT
L   D   N   P   V   N   M   I   A   D   I   E   K   G   Y   R   T   V   V   L

TTT TAG ATC TTT ACT ACT TTC CAC AAA TCC TCT TAC TCG ATA ACG TGT ATA GGG TTT TAA 240
AAA ATC TAG AAA TGA TGA AAG GTG TTT AGG AGA ATG AGC TAT TGC ACA TAT CCC AAA ATT
K   L   D   K   S   S   E   V   F   G   R   V   R   Y   R   T   Y   P   K   L

TTT GGA TTC CTC TAT TTT TTG ATA CAG ATA GGT TGG AAA TAG GGT TCC AGC AAT ATC TCT 300
AAA CCT AAG GAG ATA AAA AAC TAT GTC TAT CCA ACC TTT ATC CCA AGG TCG TTA TAG AGA
K   S   E   E   I   K   Q   Y   L   Y   T   P   F   L   T   G   A   I   D   R

TGC CCC AAA CAT ATG AAG ATA ATG TCC TTT AAA AAT GGG ACT ATC CTC TCG AAT CTC AAT 360
ACG GGG TTT GTA TAC TTC TAT TAC AGG AAA TTT TTA CCC TGA TAG GAG AGC TTA GAG TTA
A   G   F   M   H   L   Y   H   G   K   F   I   P   S   D   E   R   I   E   I

CGT ATT ACT TGT TTT CCA GCC GCT TGA TGC ATC TTC AAA ATC ACC ATT GTA GAG TAA ATT 420
GCA TAA TGA ACA AAA GGT CGG CGA ACT ACG TAG AAG TTT TAG TGG TAA CAT CTC ATT TAA
T   N   S   T   K   W   G   S   S   A   D   E   F   D   G   N   Y   L   L   N

ACG AGA CCA GCT AAG TTG TTT TGC CTG TTT TAC TTC ATC CCA TAA TAT CAT TTT TTC TTT 480
TGC TCT GGT CGA TTC AAC AAA ACG GAC AAA ATG AAG TAG GGT ATT ATA GTA AAA AAG AAA
R   S   W   S   L   Q   K   A   Q   K   V   E   D   W   L   I   M   K   E   K

GGG AGA ATG TTC ATT AGA CAT ACA TTC TAT AGA AAT AGC CGC TTG ATC TAT TTC ATA ATC 540
CCC TCT TAC AAG TAA TCT GTA TGT AAG ATA TCT TTA TCG GCG AAC TAG ATA AAG TAT TAG
P   S   H   E   N   S   M   C   E   I   S   I   A   A   Q   D   I   E   Y   D

TGT CGT TTC TAT TTT TAA TTT ATT TTT CAT CGT ATT GGT ATA CAT AGT ATT CAC TAC CTT 600
ACA GCA AAG ATA AAA ATT AAA TAA AAA GTA GCA TAA CCA TAT GTA TCA TAA GTG ATG GAA
T   T   E   I   K   L   K   N   K   M   T   N   T   Y   M   T   N   V   V   K
```

Figure 4.12 Nucleotide sequence of the insert DNA from clone pAM15 obtained from the T7 primer.
The *Sph*I restriction enzyme site used to construct the clone is underlined and this marks the start of the sequence obtained from the T7 primer. The open reading frame *orfB*, described in the text, is transcribed from the complementary strand, and is shown using the single-letter amino acid code.

obtained from the T7 primer that identified putative δ-endotoxin sequence at the edge of the construct and apparently extending past the limits of the cloned fragment. If the *Eco*RI fragment is indeed located within construct pAM15 this would indicate either that *orfA* encodes a protein of molecular weight greater than 65-kDa or that the *orfB* sequence is from a second δ-endotoxin gene immediately downstream of *orfA*. The *Eco*RI fragment was therefore selected for subcloning to facilitate sequence analysis of this region to characterise any δ-endotoxin genes present. Subcloning and sequence analysis of the 0.7 kb *Hin*dIII and the 2.5 kb *Eco*RI fragments is described in Chapter 5.

Digestion of pAM15 with *Xba*I and *Bam*HI resulted in the production of two (approximately 11 kb and 5 kb) and three (approximately 10.3 kb, 4.2 kb and 1.5 kb) restriction fragments respectively. To aid further restriction mapping of this large clone *Xba*I and *Bam*HI restriction digests of about 10 µg of pAM15 DNA were carried out and the reaction mixtures subject to agarose gel electrophoresis. The largest band from each digest was excised, the DNA repurified using a Geneclean II kit and resuspended in a volume of 20 µl. The DNA from these two bands was then used in a second round of restriction endonuclease mapping and the information used in conjunction with that obtained from mapping of whole pAM15 DNA to obtain a fuller restriction map of pAM15 (Figure 4.14).

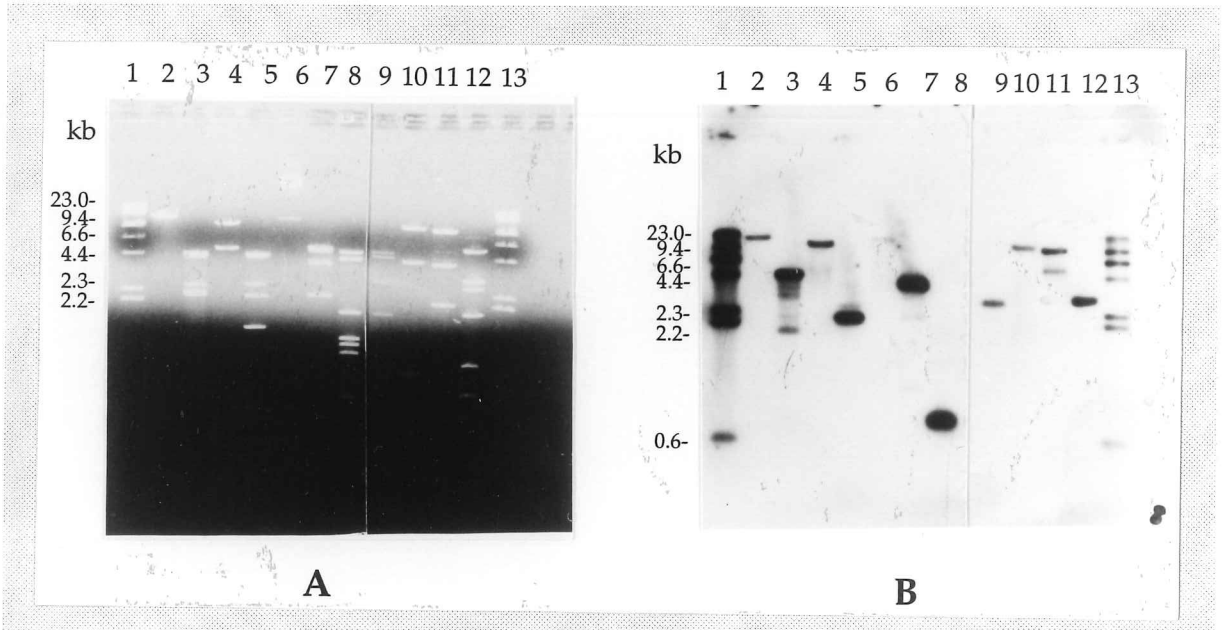


Figure 4.13 Restriction endonuclease analysis of clone pAM15. (photo no. 39E4)
Panel A: Clone pAM15 digested with the following enzymes. The size marker in lanes 1 and 13 is *Hin*dIII digested digoxygenin-labelled lambda DNA.

Lane 1	Size marker	Lane 8	*Hin*dIII
Lane 2	*Sac*I	Lane 9	*Eco*RI
Lane 3	*Kpn*I	Lane 10	*Bam*HI
Lane 4	*Xba*I	Lane 11	*Sca*I
Lane 5	*Acc*I	Lane 12	*Nde*I
Lane 6	*Hin*cII	Lane 13	Size marker
Lane 7	*Pst*I		

Panel B: Southern blot and hybridisation of Panel A with digoxygenin-labelled 17A.1.

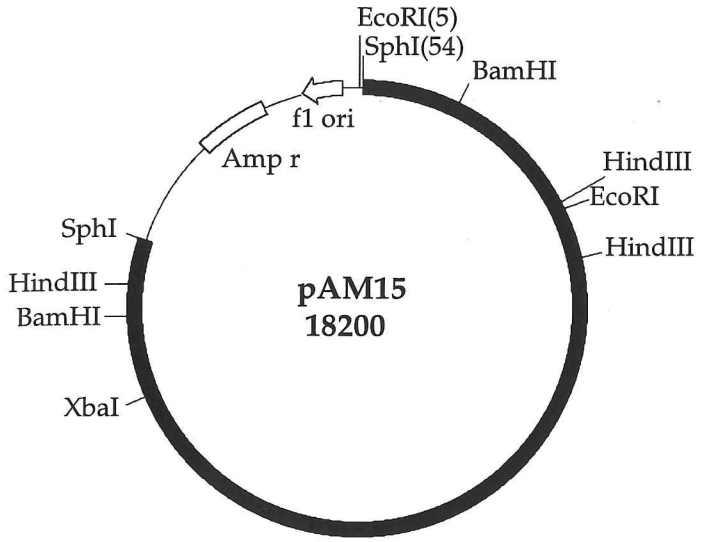


Figure 4.14 Restriction map of plasmid construct pAM15.
The 15 kb fragment (black box) from Bt ssp. *fukuokaensis* 17A is cloned into the *Sph*I site of the vector pGEM-3Zf(+). The ampicillin resistance gene is represented as Amp r and the f1 origin is also indicated. Position 5 marks the start of the pGEM-3Zf(+) polylinker. The sequence shown in Figure 4.12 starts at the site marked *Sph*I(54) and runs in a clockwise direction; *orfB* therefore runs anticlockwise.

4.5 Discussion

Plasmid DNA from Bt ssp. *fukuokaensis* 17A proved somewhat more difficult to purify than from other Bt strains. Dunn (1996) had reported the need for a particularly high SDS concentration (believed to affect both cell disruption and the folding of associated proteins) and treatment of the DNA sample with proteinase K (to help prevent degradation of the plasmid by nucleases) in the preparation of DNA from the related strain Bt ssp. *fukuokaensis* 84-I, but these additions to the protocol were found to be unnecessary if considerable care was taken when resuspending the large, sticky isopropanol pellet.

This DNA was used to identify and clone restriction fragments believed to harbour a gene encoding a novel 65-kDa Cry protein and possibly a second protein of 70-kDa. Hybridising *Hin*dIII fragments of 0.7 kb (pAM0.7) and 4.4 kb (pAM4.4), and a hybridising 15 kb (pAM15) *Sph*I fragment, were successfully cloned into the plasmid vector pGEM-3Zf(+).

0.6 kb of sequence from each end of the cloned fragments using the T7 and Sp6 universal primers identified open reading frames in constructs pAM0.7 (*orfA*) and pAM15 (*orfB*) with significant homology to *cry4A* from Bt ssp. *israelensis*. Restriction mapping carried out on pAM15 indicates that both open reading frames are transcribed in the same direction. Despite the similarity of *orfA* to the N-terminal region and *orfB* to the C-terminal region of *cry4A* it is not yet possible from the limited sequence data so far available to determine conclusively if *orfA* and *orfB* are parts of the same gene. Indeed, if the restriction

map of pAM15 is correct then *orfA* and *orfB* are separated by too great a distance to both be part of a gene encoding a protein as small as 70-kDa. Various possibilities exist:-

(i) *orfA* and *orfB* are parts of two different δ-endotoxin genes, *orfB* being located immediately downstream of and transcribed in the same direction as *orfA*.

(ii) *orfA* and *orfB* are parts of the same gene and encode a protein larger than 70-kDa (a 130-kDa protein may be indicated by the sequence similarity to the 130-kDa protein Cry4A). This may explain the presence of a faint band at about 130-kDa in some of the immunoblots (Section 3.5). The 130-kDa protein may be very poorly expressed or be C-terminally degraded to form the range of 60 to 70-kDa molecular weight bands visible on the protein profile.

(iii) *orfA* may be part of a *cry10A*-type gene. The *cry10A* gene of Bt ssp. *israelensis* encodes a protein of 84-kDa that is present in the Bti inclusion as a protein of about 70-kDa. Immediately downstream of the *cry10A* gene is a second open reading frame with significant homology to the C-terminal region of a 130-kDa Cry4-type protein (Thorne *et al.*, 1986). *cry10A* therefore appears to be derived from a 130-kDa gene that has been disrupted.

No similarity to any previously cloned δ-endotoxin genes was found in the sequence obtained from pAM4.4.

The recombinant clone pAM4.4 was found to harbour a 2.5 kb *Pst*I fragment to which the 17A.1 oligonucleotide probe hybridises. Hybridising fragments of 0.7 kb (*Hin*dIII) and 2.4 kb (*Eco*RI) were also identified in construct pAM15. These three fragments are of a useful size to facilitate subsequent sequence analysis and were selected for subcloning.

Chapter 5 describes the subcloning of the selected hybridising fragments from pAM4.4 and pAM15 and investigates the sequence data obtained for the presence of novel *cry* genes.

Five

Chapter 5

Sequence and Analysis of Clones from 17A Plasmid DNA

5.1 Introduction

The previous Chapter described the cloning of 0.7 kb (pAM0.7) and 4.4 kb (pAM4.4) *Hin*dIII fragments from Bt ssp. *fukuokaensis* 17A plasmid DNA into the vector pGEM-3Zf(+). Also, the cloning of a 15 kb (pAM15) *Sph*I fragment from the same strain. The three fragments (and consequently the three constructs) were all found to hybridise with a digoxygenin-labelled oligonucleotide probe (17A.1) derived from the N-terminal sequence of a 65-kDa protein present in the 17A parasporal inclusion. This protein was believed (from comparison of the N-terminal sequence with those of the Bti Cry proteins) to be a novel Bt δ-endotoxin. It was hoped therefore, that at least one of these constructs would harbour the gene encoding this novel protein.

Sequence obtained from the ends of each clone, using the universal priming sites present on the cloning vector, had confirmed that pAM0.7 and pAM15 did indeed contain open reading frames (*orfA* and *orfB*, respectively) with homology to other Bt crystal protein genes (as determined by DNA and protein BLAST database searches), most notably Cry4A from Bt ssp. *israelensis*. Analysis of sequence obtained from the ends of clone pAM4.4 did not however, show any significant homology to any previously identified δ-endotoxin gene.

0.6 kb of sequence data had been obtained from each of the two universal primers for each clone. This was sufficient to provide unambiguous sequence information for the whole of pAM0.7. However, about 3.2 kb of pAM4.4 and some 13.8 kb of pAM15 still remained unsequenced. Before proceeding further with sequencing of these clones, it was important to develop an overall stategy that took into consideration the overall size of the DNA to be sequenced, the accuracy of the sequence required and the facilities available. Due to the large size of the remaining constructs compared to the size of a gene that would be needed to encode a 65-kDa protein (about 2.1 kb), it was decided to construct subclones containing smaller regions of hybridising DNA. It was hoped that these subcloned regions would facilitate sequencing of the putative *cry* gene and provide rapid identification of the N-terminus. Restriction analysis of pAM4.4 had identified a 2.5 kb *Pst*I fragment that hybridised the 17A.1 probe. Hybridising fragments of 0.7 kb (*Hin*dIII) and 2.5 kb (*Eco*RI) were also identified in construct pAM15. These three fragments were therefore selected and their subcloning and sequence analysis is described in this Chapter.

To sequence the subcloned DNA three main routes were available. The first of these is a random approach ('shotgun' sequencing), in which sequence can be collected from subclones containing random fragments of the target DNA. In this method no attempt is made to determine where these subclones map in the target DNA or which strand of DNA is being sequenced. Instead the accumulated data are processed by computer (Staden, 1984) and arranged into contiguous sequence (Anderson *et al.*, 1981; Baer *et al.*, 1984). Alternatively, one of two directed systematic methods can be used. The fragment can be sequenced by creating a set of nested deletion mutants that begin at a common point and penetrate various distances into the target sequence bringing progressively more remote regions of DNA into range for sequencing by the universal primer. Or, the DNA fragment can be sequenced in a step-wise fashion using oligonucleotides designed to hybridise to the ends of sequence obtained from a previous reaction as primers in a subsequent reaction.

Initially the method of choice for this work was to create a set of unidirectional nested deletions whose overlapping sequences could then be assembled to obtain a complete contiguous fragment of novel DNA sequence.

5.2 Sequencing Strategy for Construct pAM15

2.5 kb *Eco*RI and 0.7 kb *Hin*dIII hybridising fragments were selected for subcloning into the vector pGEM-3Zf(+). The 2.5 kb *Eco*RI fragment was selected because it was large enough to contain most, if not all, of the putative *cry* gene, yet was still a manageable size for sequencing. The 0.7 kb *Hin*dIII fragment was to be sequenced in order to confirm that this same fragment was present in construct pAM0.7.

In order to subclone the *Eco*RI fragment, about 50 µg of *Eco*RI digested pAM15 was run on a 0.5% agarose gel in 1x TAE buffer, the 2.5 kb band excised, repurified using a Geneclean II kit and resuspended in a volume of 20 µl. This insert fragment was ligated to *Eco*RI-linearised, dephosphorylated pGEM-3Zf(+) (in an approximately 1:3 molar ratio of insert to vector, as determined by agarose gel electrophoresis) and transformed into 50 µl of *E. coli* JM109 under the conditions described in Chapter 4 for the cloning of the initial fragments. This procedure was repeated using *Hin*dIII linearised pGEM-3Zf(+) in order to subclone the 0.7 kb fragment isolated from a *Hin*dIII digest of pAM15 also by gel purification. Subclones of both the 2.5 kb *Eco*RI fragment (pAMscE) and the 0.7kb *Hin*dIII fragment (pAMscH) were obtained and their identity confirmed by Southern blotting and positive hybridisation with the probe 17A.1.

Sequence obtained from each end of the subcloned DNA using the T7 and Sp6 primers confirmed that the 0.7 kb fragment was the same as that identified in construct pAM0.7 (see Figure 4.7). It also confirmed the pAM15 restriction analysis data that

suggested that the 2.5 kb *Eco*RI fragment would start with the first codon of the open reading frame *orfA* identified in pAM0.7. The sequence obtained from pAMscE also revealed that an almost complete pGEM-3Zf(+) polylinker was present on either side of the subcloned DNA (see Figure 5.1). For this to occur the first *Eco*RI site must have been in the Bt DNA (*Eco*RI(B) in Figure 5.1) and the second within the pGEM polylinker, indicating that it had been cloned from one extreme end of the Bt DNA contained in pAM15 (Figure 5.1). The presence of duplicate polylinkers meant that nested deletions could not be made from this clone to facilitate sequencing as had initially been intended. This method (Henikoff, 1984 and 1987) exploits the ability of exonuclease III to catalyse the step-wise removal of 5' mononucleotides from the recessed or blunt 3' hydroxyl termini of double-stranded DNA (Weiss, 1976), and its inability to do the same to protruding 3' termini. Using a time course a set of unidirectional deletions of the target DNA is therefore produced. Unique sites within the vector polylinker are therefore required to produce both a nuclease-susceptible and a nuclease-resistent end.

Sequencing of pAMscE was therefore carried out by "gene walking" using oligonucleotides derived from the previous sequencing reaction as primers for a subsequent round of sequencing, as described in the introduction to this Chapter. This step-wise method of sequencing generated about 500-600 bp of unambiguous sequence from each end of pAMscE per reaction and was repeated until overlapping sequence had been obtained for both strands of the clone. These sequences were then compiled alongside that obtained from construct pAM0.7 using the 'bestfit' program from GCG (Section 2.1.6) to obtain a 3.3 kb contiguous fragment of novel DNA sequence.

5.3 Sequence Analysis

The 3.3 kb of sequence data obtained from pAM0.7 and pAMscE is shown between nucleotides 272 and 3621 in Figure 5.10. When translated, two open reading frames were identified. The first open reading frame included the *orfA* identified in pAM0.7 and ran for 1857 bp encoding a 699 amino acid protein of predicted weight 79,063 Da. At nearly 80-kDa the predicted molecular weight of this protein obviously exceeds that of the 65-kDa protein identified in the 17A inclusion. From the sequence information obtained so far it is difficult to determine if the cloned gene is indeed that encoding the 65-kDa protein. It is possible that an unstable 80-kDa protein is degraded to a stable 65-kDa end product as is the case with Cry10A (Purcell, 1997).

BLAST database searches were carried out on both the DNA and predicted protein sequence of this first open reading frame. Significant levels of identity and similarity were

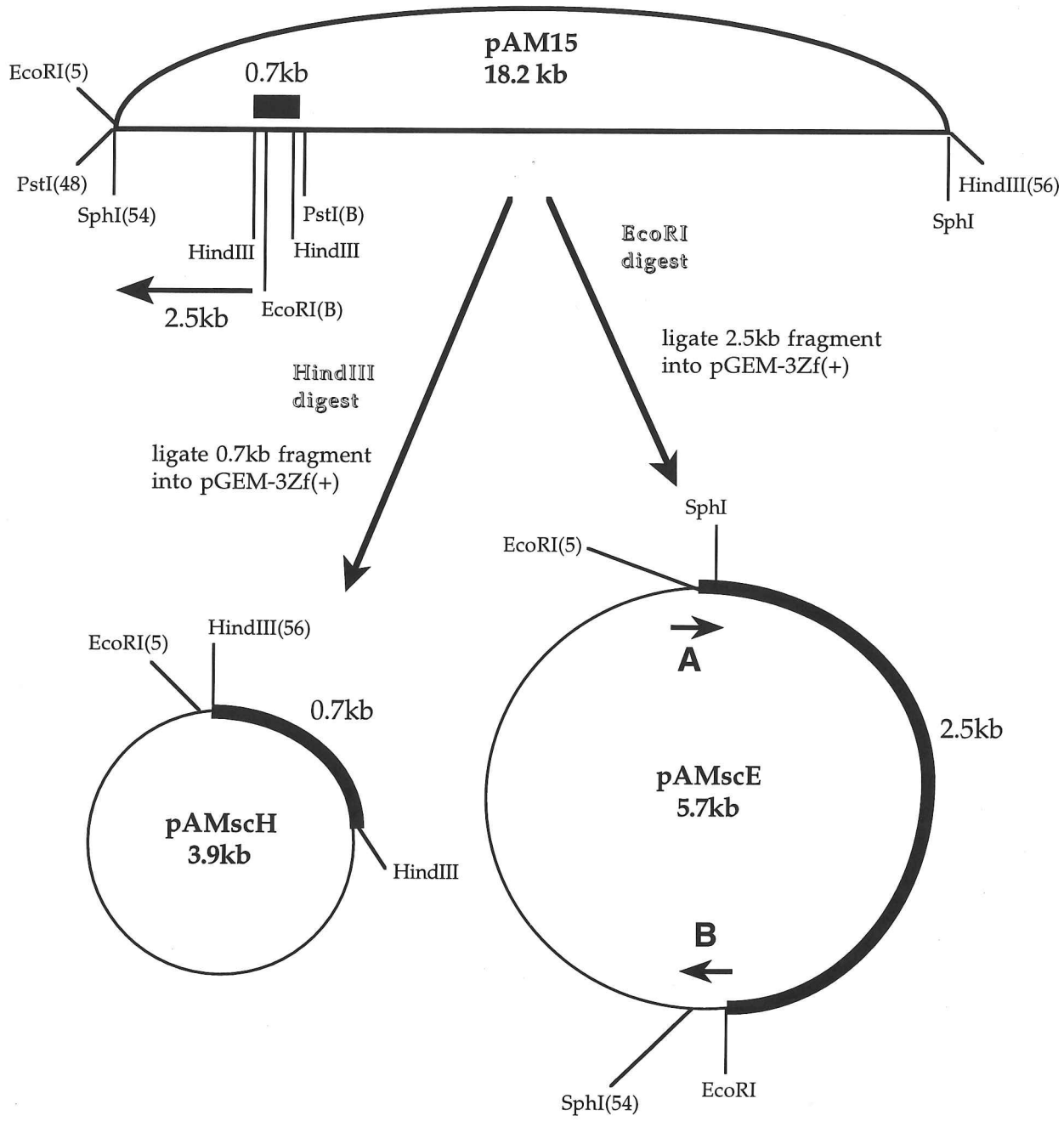


Figure 5.1 Subcloning of a 0.7 kb *Hin*dIII fragment and a 2.5 kb *Eco*RI fragment from pAM15.
Within the pAM15 clone the position of the 0.7 kb *Hin*dIII fragment is indicated by a labelled black box. The position of the 2.5 kb *Eco*RI fragment is indicated by a labelled arrow that also indicates the direction of transcription of the open reading frames so far identified within this clone (*orfA* and *orfB*). Arrows A and B denote the positions of the two almost complete pGEM-3Zf(+) multiple cloning site polylinkers present in construct pAMscE. Polylinker A is derived from construct pAM15 and polylinker B from the new vector molecule. The numbers in brackets after the restriction sites indicate their positions in the original vector before addition of the cloned fragment.

found to several Bt crystal protein genes, notably, as had been predicted from the N-terminal amino acid sequence, the 130-kDa Cry4A (33%) and Cry4B (31%) proteins and the 66-kDa Cry10A (31%) protein, all from Bt ssp. *israelensis*. Figure 5.2 shows the percentage amino acid sequence similarity between the Bti Cry proteins and the putative 65-kDa δ-endotoxin from 17A (hereafter called AM1) calculated using the Clustal method with PAM250 residue weight table. It can be seen from this data that AM1 shows less similarity to Cry11A (13.5%) than to Cry4A, Cry4B and Cry10A (all above 20%). This information, combined with the BLAST alignment data, where AM1 was aligned to the N-terminal region of Cry4A and 4B, indicates that the 65-kDa AM1 is likely to be a truncated Cry4A/4B-type protein much like Cry10A, rather than a Cry11A-type protein, despite the similarity in length and N-terminal sequence. In the old nomenclature Cry4A, Cry4B, Cry10A and Cry11A were named CryIVA, CryIVB, CryIVC and CryIVD, respectively. The reclassification of Cry10A (IVC) and Cry11A (IVD) into new groups away from Cry4A and 4B reflects the sequence divergence of these proteins, both from 4A and 4B and from each other, as shown in Figure 5.2.

Percentage Similarity

	AM1	Cry4A	Cry4B	Cry10A	Cry11A	
AM1		22.9	23.5	23.3	13.5	AM1
Cry4A	65.3		52.5	27.1	10.7	Cry4A
Cry4B	66.7	40.3		23.6	11.0	Cry4B
Cry10A	67.6	64.2	68.6		11.5	Cry10A
Cry11A	80.3	80.0	79.8	83.6		Cry11A
	AM1	Cry4A	Cry4B	Cry10A	Cry11A	

Percentage Divergence

Figure 5.2 Protein sequence relationships between AM1 and the Cry proteins from Bt ssp. *israelensis* using a PAM250 residue weight table.

The Clustal alignment program predicted levels of similarity between AM1 and the Bti Cry sequences about 10% lower than had been predicted by the pair-wise alignments carried out using the 'gap' and 'bestfit' programs from GCG. When the alignments were examined it was found that large gaps had been introduced towards the end of the AM1 sequence in an attempt to align it to the whole Cry4A and Cry4B sequence, ignoring what appeared to be well-conserved motifs present in the GCG alignments. Pair-wise alignments of AM1 with Cry4A and B and Cry10A were also carried out using the Lipman-Pearson

Protein alignment program (also part of the Megalign package) and these results agreed largely with those obtained from the GCG programs in that all obvious sequence motifs were recognised. Attempts to align AM1 with just the first 700 amino acid residues of Cry4A and Cry4B did little to improve the Clustal alignments and the percentage similarities remained about 10% lower than had originally been predicted. The problem seemed to be in achieving multiple alignments of what are fairly divergent sequences.

Eventually the stacked Clustal alignment of AM1, Cry4A, Cry4B and Cry10A was edited by hand using the Lipman-Pearson Protein alignments to identify and realign conserved motifs and residues. For example, the strongly conserved motif, ID-IE--P, towards the end of the final sequence alignment shown in Figure 5.4 (underlined in this Figure) was not aligned in the original Clustal alignment. This process considerably improved the calculated similarity (shown in Figure 5.3) not only between AM1 and the Bti Cry protein sequences but also between the Bti sequences themselves, indicating that the changes made reflected genuine conserved sequence patterns.

Percentage Similarity

	AM1	Cry4A	Cry4B	Cry10A	
AM1		29.9	28.7	30.4	AM1
Cry4A	64.1		54.6	32.9	Cry4A
Cry4B	65.9	39.3		29.2	Cry4B
Cry10A	66.6	65.5	67.4		Cry10A
	AM1	Cry4A	Cry4B	Cry10A	

Percentage Divergence

Figure 5.3 Protein sequence relationships between AM1 and the highest scoring matches from Figure 5.2 using a PAM250 residue weight table.
The alignment was then edited by hand to realign highly conserved motifs as described in Figure 5.4.

Figure 5.4 Clustal alignment of the amino acid sequence of AM1 with the N-terminal regions of Cry4A and Cry4B and the complete Cry10A protein from Bt ssp. *israelensis* (continued over page).

The original multi-alignment was generated using a PAM250 residue weight table. This was then realigned by hand with reference to pairwise Lipman-Pearson Protein alignments of AM1 with each of the above proteins to maintain conserved motifs (such as the underlined motif towards the end of the alignment). Conserved or semiconserved residues are shown in bold and marked with an asterisk. Residues that are conserved between AM1 and all but one other sequence are marked with an asterisk in plain text.

```
AM1       MNSYQNKNEYEILDASRNNSNMSNRYPRYPLANDPQACMQNTNYKDWLATC--N---GT
Cry4A     MNPYQNKNEYETLNASQKKLNISNNYTRYPIENSPKQLLQSTNYKDWLNMCQQNQQYGG
Cry4B     MNS----------------------GYPLANDLQGSMKNTNYKDWLAMCENNQQYGV
Cry10A    MNPYQNKNEYEIFNAPSNGFSKSNNYSRYPLANKPNQPLKNTNYKDWLNVCQDNQQYGN
          ** ********    *        ** * ****** *     ********  *  *   *

AM1       TVPFSNPSQLLKV---GGIVVQRVLGGIATFFPGIGPLVPFLTFFASLLWPSGSSG-NDIW
Cry4A     DFETFIDSGEL---SAYTIVVGTVLTGFG-FTTPLGLALIGFGTLIPVLFPAQDQS-NT-W
Cry4B     NPAA-INSSSV---STALKDAGAILKFVNPPA---GSVLTVLSAVLPILWPTNTPTPERVW
Cry10A    NAGNFASSETIVGVSAGIIVVGTMLGAFAAPVLAAGIISFGTLL--PIFWQ-GSDPAN-VW
                *          ***   *          *            ***      *  *

AM1       EKLMKEVADLIQQELTTYTIDKATAELGGLKELLDSYNRAFASWEVGNATPGLVKGYIES-
Cry4A     SDFITQTKNIIKKEIASTYISNANKILNRSFNVISTYHNHLKTWENNPNPQNTQDVRTQIQ
Cry4B     NDFMTNTGNLIDQTVTAYVRTDANAKMTVVKDYLDQYTTKFNTWKREPNNQSYRTAVITQF
Cry10A    QDLLNIGGRPI-QIEDKNIINVLTSIVTPIKNQLDKYQEFFDKWEPARTHANAK-AVHD--
                  * **     *  *       *  ** *   *  **

AM1       -L-HRKFVDKVISGFTQP-------GYEKIL-LPSYAIAANFHLLVLRDIEIYGERLGFS
Cry4A     -LVHYHFQNVIPELVNSCPPNPSDCDYYNILVLSSYAQAANLHLTVLNQAVKFEAYLKNN
Cry4B     NLTSAKLRETAVYF--SNLV-----GYELL-LLPIYAQVANFNLLLIRDGPHKCTRMVYA
Cry10A    -LFTTLEPIIDKDLDMLKNNA----SYRIPT-LPAYAQIATWHLNLLKHAATYYNIWLQN
           *                        *     ** **  ** **  *

AM1       QV-DRNF-------YNCELKFFMAKYTNYCVDTYNKGLASEKEKG---------WVPFHR
Cry4A     RQFDYLEPLPTAIDYYPVLTKAIEDYTNYCVTTYKKGLNLIKTTPDSNLDGNINWNTYNT
CRY4B     RSCDQ--------LYNT-MVQYTKEYIAHSITWYNKGLDVLRNKSNGQ------WITFND
CRY10A    QGIN-PSTFNSSNYYQGYLKRKIQEYTDYCIQTYNAGLTMIRTNTNA------TWNMYNT
             *          *   *      ** **  ******                 *

AM1       YRREMTLAVLDIIALFPLYDARLYPATD-NKEMPVKSELTREIYSD-VINSDRFGIVPP
Cry4A     YRTKMTTAVLDVVALFPNYDVGKYPIG-------VQSELTREIYQVL--NFEE-SPYKY
Cry4B     YKREMTIQVLDILALFASYDPRRYPADKIDNTKLSKTEFTREIYTALV----E-SP-SS
Cry10A    YRLEMTLTVLDLIAIFPNYDPEKYPIG-------VKSELIREVYTN-V-NSDTFR----
          ** ***  ***  **** **   **          **********   * *

AM1       YNYSQNEERYTRPPHLFTWLRGLDFVTNVLTSGTYAYRWRVLTGCQNKYSYTRGNGTITG
Cry4A     YDFQYQEDSLTRRPHLFTWLDSLNFYEKA---QTTPNN--FFTSHYNMFHYTLDNISQKS
Cry4B     KSIAALEAALTRDVHLFTWLKRVDFWTNT---IYQDLR--FLSANKIGFSYTNSSAMQES
Cry10A    -TITELENGLTRNPTLFTWINQGRFYTRNSRDILDPYDIFSFTGNQMAFTHTND---DRN
             *   ** *******    * *                 *       **
AM1       PFRGYPVESGGSTSNITIEEGSYIYNLLPRSFEYISPWYFTT-NIAMITFLLTNNNSSTEK
Cry4A     SVFG-NHNVTDKLKSLGLATNIYIF-LLN-VISLDNKYLNDYNNISKMDFFITNGTRLLEK
Cry4B     GIYG-SSGLVQ----------IYLI-KFN--LILIVIKLLSQILAPPLIELQKWISTKFDG
Cry10A    IIWG---AVHGNIISQDTSKVFPFYRNKPIDKVEIVRHREYSDIIYEMIFFSNSSEVFR--
             *                  *          *        *    *
```

Figure 5.4 Clustal alignment of the amino acid sequence of AM1 with the N-terminal regions of Cry4A and Cry4B and the complete Cry10A protein from Bt ssp. *israelensis* (continued over page).

The original multi-alignment was generated using a PAM250 residue weight table. This was then realigned by hand with reference to pairwise Lipman-Pearson Protein alignments of AM1 with each of the above proteins to maintain conserved motifs (such as the underlined motif towards the end of the alignment). Conserved or semiconserved residues are shown in bold and marked with an asterisk. Residues that are conserved between AM1 and all but one other sequence are marked with an asterisk in plain text.

```
AM1       -VY-GSLVNQPNLPIVQTDFDFLTNKEVTGS----PTYNNYNHILSYMLLGYDWNQTGGIG
Cry4A     ELTAGSGQITYDVNKNIFGLPILKRRENQGNPTLFPTYDNYSHILSFIK-SL--SIPATYK
Cry4B     TLASYNSNITPTPEGLRTTFFGFSTNENTPNQ---PTVNDYTHILSYIK-TD--VI--DYN
Cry10A    --YSSNSTIENNYKRT--DSYMIPKQTWKN--------EEYGHTLSYIK-TDNYIFSVVRE
                *                      *        **    * *****

AM1        THGYSFAFTHSSVDPYNTIASDKITQIPAVKGHSLQNGI-VVGGPGHTGGDLVNMGYA
Cry4A      TQVYTFAWTHSSVDPKNTIYTHLTTQIPAVKANSLGTASKVVQGPGHTGGDLIDFK--
Cry4B      SNRVSFAWTHNIVDPNNQIYTDAITQVPAVKSNFLNATARVIKGPGHTGGDLVALTSN
Cry10A     RRRVAFSWTHTSVDFQNTIDLDNITQIHALKALKVSSDSKIVKGPGHTGGDLVILKDS
               *** ******* ***   * ********    *      ** **********

AM1        STLS-----MNC---YFSQPL-NYKMRIRYSTSFYTPF-YISSPHRSGVVSISLFQTNYVK
Cry4A      -----DHFKITCQHSNFQQ---SYFIRIRYASNGSA--NTRAVINLSIPGVAELGMALNPT
Cry4B      GTLSGGRMEIQCKTSIFNDPTRSYGLRIRYAANSPIVIECDHMYYKEFLEEQRLVQNYVSR
Cry10A     ---------MDFRVRFLKNVSRQYQVRIRYATNAP-----KTTVFLTGIDTISVELPSTTS
                     *    *      *  *******                        *

AM1        VDG-KNIEVDLMDPRAFQIIEVPVEFQATSSGYANLIFT---ANASYGVYIDKIEFIPSN
CryIVA     FSGTDYTNLKYKDFQYLEFSNEV---KFAPNQNISLVFNRSDVYTNTTVLIDKIEFLPIT
CryIVB     PNNIIPTDLKYEEFRYKDPNDAIVPMRLSSNQLITIAIQPLNMTSNNQVIIDRIEIIPIT
CryIVC     RQNPNATDLTYADFGYVTFPRTVPNKTFEGEDTLLMTLY-GTPNHSYNIYIDKIEFIPIT
               *    *                                      * ********

AM1        IQIDKCTKCQFEEKVCTCRCEGVQSLETEKEIVNSLFIN
Cry4A      RSIREDREKQKLETVQQIINTFYANPIKNTLQSELTDYDIDQAANL
Cry4B      QSVLDETENQNLESEREVVNALFTNDAKDALNIGTTDYDIDQAANL
Cry10A     QSVLDYTEKQNIEKTQKIVNDLFVN
             *   *  *  *           *
```

Figure 5.4 continued.

It can be seen from Figure 5.4 that although some regions of the four sequences share little similarity there are residues and motifs that are clearly conserved along the entire length of the alignment. This perhaps indicates that all four proteins are derived from a common ancestor. It is particularly interesting to note that AM1 is no more closely related to Cry10A than it is to Cry4A or B. This relationship is more clearly demonstrated in the phlyogentic tree shown in Figure 5.5. It might have been expected from the similarity in size of AM1 and Cry10A (both about 70-kDa) compared to that of the larger Cry4A and B proteins (130-kDa) that AM1 and Cry10A had diverged relatively recently and might share a level of similarity such as is seen between Cry4A and B (see Figure 5.3). Such a level of similarity would put AM1 and Cry10A into the same group for the new nomenclature and suggest that AM1 should be named Cry10B. However, Figure 5.5 clearly shows that this is not the case and AM1 and Cry10A are no more closely related than are Cry10A and Cry4A. This suggests that AM1 can be considered as a novel Bt δ-endotoxin and given a nomenclature that reflects this.

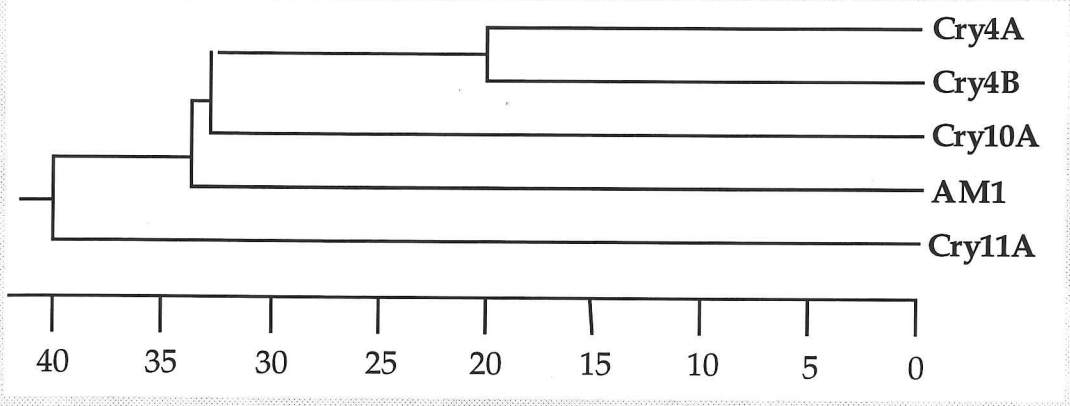


Figure 5.5 Phylogenetic tree of alignment from Figure 5.3.
The length of each pair of branches represents the distance between sequence pairs. The scale beneath the tree measures the distance between sequences; units indicate the number of substitition events.

Analysis of the sequence upstream of *am1* revealed a potential Shine-Dalgarno ribosome binding sequence (RBS; Shine & Dalgarno, 1974) GGGAG 13 base-pairs upstream of the putative ATG initiation codon (Figure 5.10). This region was also analysed for the presence of promoter regions which have been identified for other Bt δ-endotoxins. As has already been discussed in the Introduction to this Thesis, δ-endotoxin synthesis is temporally regulated at the transcription level by the activation of sporulation-specific sigma factors which, by binding to RNA polymerase, recognise the δ-endotoxin gene promoters. The *cry1Aa* gene is a typical example. Two overlapping start sites have been mapped, defining two overlapping promoters (BtI and BtII) which are used sequentially (Wong *et al.*, 1983) in transcription of this gene. BtI is active between about t_2 and t_6 of sporulation, and BtII from t_5 onwards (where t_n is n hours after the end of the exponential phase). The RNA polymerase recognising these sequences during sporulation has been purified (Brown & Whiteley, 1988) and two proteins of 35-kDa and 28-kDa (σ^{35} and σ^{28}) which form part of this polymerase have been cloned and chemically characterised. Many other *cry* genes including *cry4A, cry4B, cry11A* and *cyt1A* from Bt ssp. *israelensis* have promoter regions containing consensus regions for either BtI and BtII or BtI alone (Dervyn *et al.*, 1995; Yoshisue *et al.*, 1993a and 1993b).

Putative -35 and -10 promoter regions were identified upstream of *am1* (Figure 5.6) with similarity to the σ^{35} RNA polymerase (Brown & Whiteley, 1988, 1990). No obvious homology to the σ^{28} promoter region was found.

```
     -35                    -10
TGTGCATTTTTTCATAAGATGAGT-CATATGTTTTAAATT        cry1Ac[1]
TATGAATATTTTCATAAGCTGAAC-CATATGATTTAAACT        cry1Ba[2]
CAGGAACATTTTGAAATTTCCCG--CATAGAATACTGAA         cry15Aa[3]
   GCATATACCGTTTACTCCCCA-CATAGAATGTGCAGA        cry2Aa[4]
CAGGCATCTTTCGAACTATAGCG--CATAGAATACTA           cyt1A[5]
CAGGAATAGATTATTTAAAATTACGAATACTTTAAAATG         cry4A[6]
ATGGAATAATTATATTGGTACAGA-AATATGATTGGGATTA       cry4B[7]
GTAGGGAAATGGATACCTTTGTG--CAAACAGGAGGGGTTT       cry10A[8]
CATGCATCGTTTTTATACAAGTAA-CATATATTTGTTATG        cry11A[9]
   GCATATACCTATTTTTTTAA--CATACACATAAAAAGA       84-Iorf1[10]
TACGAATATATAACAAGACACCTAGCATATACTTTGGGT         17Aam1
   KMATATT                 CATACA-T             consensus
```

Figure 5.6 Promoter Alignment.
Promoters of sporulation-specific genes. Alignments of the promoters of sporulation dependent ICP genes adapted from Baum and Malvar (1995) with the addition of promoter regions from *cry10A*, the Bt ssp. *fukuokaensis* 84-I *orf*1 gene (denoted 84-Iorf1) and the Bt ssp. *fukuokaensis* 17A *am1* gene (denoted 17Aam1). Proposed -10 and -35 regions are in bold. The consensus sequence is that reported for *B. subtilis* (Moran, 1993): K=G or T; M=C or A. References cited:

1. Wong *et al.*, 1983
2. Brizzard *et al.*, 1991
3. Brown, 1993
4. Widner & Whiteley, 1989
5. Ward *et al.*, 1986
6. Yoshisue *et al.*, 1993a
7. Yoshisue *et al.*, 1993b
8. Thorne *et al.*, 1986
9. Dervyn *et al.*, 1995
10. Dunn, 1996

A second open reading frame was also identified with a potential ATG start site just 70 base-pairs downstream of the *am1* TAA stop codon (position 2970) and orientated in the same direction. The 738 bp of this second open reading frame spans the rest of construct pAMscE and is shown between nucleotides 3039 and 3621 in Figure 5.10. Of the sequence available from pAMscE this second *orf* (hereafter referred to as *am2*) would encode a 246 amino acid protein with a predicted molecular weight of 28,307 Da. DNA and protein BLAST database searches revealed about 50% similarity to the C-terminal regions of the Bti Cry4A and Cry4B proteins, immediately suggesting that *am2* may have once formed a continuous open reading frame with *am1* encoding a protein of about 130-kDa. The intergenic, non-coding region between the two open reading frames was sequenced several times to ensure that the interruption was not due to sequencing error, but it was found that at least two stop codons and a frame-shift prevented the two *orfs* from encoding a single polypeptide chain.

The identification of a potential C-terminal region to AM1 raised the possibility that the very faint 130-kDa band visible in the immunoblots (Figure 3.11) could be the result of very low level read through from the promoter region of *am1* into the second open reading frame *am2*. This is however unlikely, given the discovery that *am2* is in a different reading frame from *am1*.

DNA sequence upstream of *am2* was analysed for control regions such as had been found upstream of *am1*, but no sequence was found with any real similarity to the -10 and -35 box region of Bt promoters either in front of the first ATG codon of *am2* or in front of the start of the open reading frame seven amino acids upstream. However, a potential RBS site GGAGG was identified 56 base-pairs upstream of the start of *am2* making it possible that *am1* and *am2* could form a polycistronic RNA when transcribed. The discovery of a second open reading frame encoding an independent C-terminal region is of particular interest with reference to the Cry10A protein of Bti.

The Cry10A δ-endotoxin gene was originally cloned by Thorne *et al.* (1986) from a bacteriophage library of Bt ssp. *israelensis* plasmid DNA. The gene product of this clone accumulated in *B. subtilis* as a protein with a relative molecular mass of about 58-kDa and was toxic to mosquito larvae but had no haemolytic activity. Sequencing of this clone revealed the presence of two large open reading frames for translation, *ORF1* and *ORF2*, which were oriented in the same direction. *ORF1* encoded a protein of about 72-kDa. The discrepancy in the size of the gene sequence and the expressed protein product was believed to be the result of proteolytic processing, because polypeptides as large as 72-kDa were expressed from the cloned insert when gene expression was carried out *in vitro* with an *E. coli* transcription/translation system. *ORF2* could have encoded a protein of at least 28-kDa with the C-terminal region extending beyond the boundary of the original cloned fragment. This also appears to be the case for clone pAM15. The remaining region of *ORF2* has never been cloned and so only the first 238 amino acids have ever been sequenced. The potential start for *ORF2* was identified just 64 base-pairs downstream of the TAA stop codon of *ORF1*. Removal of two stop codons in the intergenic regions between these open reading frames would form ORF1, the intergenic region and ORF2 into a single polypeptide chain.

Further work has since been carried out on Cry10A by two researchers in this laboratory. The Cry10A protein was "reconstructed" by attaching the C-terminal region of the *cry4B* gene to *cry10A* (Zhang, personal communication, as described in Purcell, 1997). This construct was then expressed in an acrystalliferous strain of Bt and produced inclusions containing a polypeptide of about 130-kDa. Contrary to the work carried out by Thorne *et al.* (1986), the *cry10A* product was found to be a 65-kDa protein non-toxic to *Aedes aegypti* larvae (Purcell, 1997). However, the Cry10A/4B chimera was found to be weakly mosquitocidal (Purcell, 1997). An investigation into the polypeptide content of the Bti native inclusion (Purcell, 1997) also revealed the presence of a 60-kDa band, the N-terminal sequence of which matched the start of the *cry10A ORF2* sequence from position 3035 as determined by Thorne *et al.* (1986). This indicated that *ORF2* was independently expressed as a 60-kDa protein at levels that were determined to be about one third that of Cry10A (Pucell, 1997). Sequence analysis of the intergenic region between *ORF1* and *ORF2*

DNA sequence upstream of *am2* was analysed for control regions such as had been found upstream of *am1*, but no sequence was found with any real similarity to the -10 and -35 box region of Bt promoters either in front of the first ATG codon of *am2* or in front of the start of the open reading frame seven amino acids upstream. However, a potential RBS site GGAGG was identified 56 base-pairs upstream of the start of *am2* making it possible that *am1* and *am2* could form a polycistronic RNA when transcribed. The discovery of a second open reading frame encoding an independent C-terminal region is of particular interest with reference to the Cry10A protein of Bti.

The Cry10A δ-endotoxin gene was originally cloned by Thorne *et al.* (1986) from a bacteriophage library of Bt ssp. *israelensis* plasmid DNA. The gene product of this clone accumulated in *B. subtilis* as a protein with a relative molecular mass of about 58-kDa and was toxic to mosquito larvae but had no haemolytic activity. Sequencing of this clone revealed the presence of two large open reading frames for translation, *ORF1* and *ORF2*, which were oriented in the same direction. *ORF1* encoded a protein of about 72-kDa. The discrepancy in the size of the gene sequence and the expressed protein product was believed to be the result of proteolytic processing, because polypeptides as large as 72-kDa were expressed from the cloned insert when gene expression was carried out *in vitro* with an *E. coli* transcription/translation system. *ORF2* could have encoded a protein of at least 28-kDa with the C-terminal region extending beyond the boundary of the original cloned fragment. This also appears to be the case for clone pAM15. The remaining region of *ORF2* has never been cloned and so only the first 238 amino acids have ever been sequenced. The potential start for *ORF2* was identified just 64 base-pairs downstream of the TAA stop codon of *ORF1*. Removal of two stop codons in the intergenic regions between these open reading frames would form ORF1, the intergenic region and ORF2 into a single polypeptide chain.

Further work has since been carried out on Cry10A by two researchers in this laboratory. The Cry10A protein was "reconstructed" by attaching the C-terminal region of the *cry4B* gene to *cry10A* (Zhang, personal communication, as described in Purcell, 1997). This construct was then expressed in an acrystalliferous strain of Bt and produced inclusions containing a polypeptide of about 130-kDa. Contrary to the work carried out by Thorne *et al.* (1986), the *cry10A* product was found to be a 65-kDa protein non-toxic to *Aedes aegypti* larvae (Purcell, 1997). However, the Cry10A/4B chimera was found to be weakly mosquitocidal (Purcell, 1997). An investigation into the polypeptide content of the Bti native inclusion (Purcell, 1997) also revealed the presence of a 60-kDa band, the N-terminal sequence of which matched the start of the *cry10A ORF2* sequence from position 3035 as determined by Thorne *et al.* (1986). This indicated that *ORF2* was independently expressed as a 60-kDa protein at levels that were determined to be about one third that of Cry10A (Pucell, 1997). Sequence analysis of the intergenic region between *ORF1* and *ORF2*

does not reveal any similarities to recognised Bt promoters; however, a potential ribosome binding site does exist (Thorne *et al.*, 1986) raising the possibility that one-third of the ribosomes that recognise the *ORF1* promotor may be able to reattach at this RBS and transcribe *ORF2*.

Before thorough analysis of *am2* could be carried out, it was necessary to clone and sequence the remainder of this open reading frame. Initial BLAST alignments of *am2* protein sequence with Cry4A and Cry4B indicated that DNA encoding about 300 amino acids was missing from the end of *am2* in construct pAM15. Restriction analysis of clone pAM15 had already been carried out as described in Chapter 4. From this information and the sequence obtained from pAMscE, a *Pst*I site was identified about 0.2 kb upstream of the predicted start of *am1* (PstI(B) in Figure 5.1); no further *Pst*I sites were identified downstream of this site in pAMscE. Previous experiments (see Figure 4.5) had already identified two N-terminal hybridising bands (10 kb and 2.1 kb) from a *Pst*I digest of total 17A plasmid DNA, only the larger of which could contain all 3.9 kb of DNA present in pAM15. PstI(B) is situated just 0.2 kb upstream of *am1* and the combined length of DNA needed to encode *am1*, *am2* and the intergenic region is only about 4 kb; this 10 kb *Pst*I fragment must therefore harbour all of the open reading frame encoding the putative AM2 protein.

In order to clone the 10 kb *Pst*I fragment, 100 µg of caesium chloride-gradient purified Bt ssp. *fukuokaensis* 17A DNA was digested with *Pst*I in a total volume of 200 µl for eight hours. The digested DNA was then electrophoresed on a 0.5% agarose gel in 1x TAE at 40 volts for eighteen hours. Five segments from the digest in the 8-12 kb region were excised along with five segments in the 1.5-3 kb region in an attempt to clone the 2.1 kb hybridising fragment so that this sequence could also be analysed. The DNA from each segment was repurified using a Geneclean II kit (Section 2.2.5) and resuspended in a volume of 20 µl. Transformation of the ligation mixes for these ten segments (see Chapter 4 for details) gave 50-100 colonies per plate of which about 50% were white. Rounds of Southern blot hybridisation with batch and then individual DNA mini-preparations (See Chapter 4 for details) identified one positive clone containing the 10 kb fragment (pAM10). No clones were identified containing the 2.1 kb fragment.

Restriction digest analysis was carried out on clone pAM10 as described for construct pAM15 in Chapter 4. A restriction map is shown in Figure 5.7. *Xba*I digestion generated four fragments of approximately 7.2 kb, 4.2 kb, 1.2 kb and 1 kb. The 7.2 kb fragment hybridised to digoxygenein- labelled 17A.1 indicating the presence of the *am1* N-terminus. These four fragments were selected as suitable for generating subclones of pAM10 to provide smaller more manageable fragments to aid the sequenceing of *am2*.

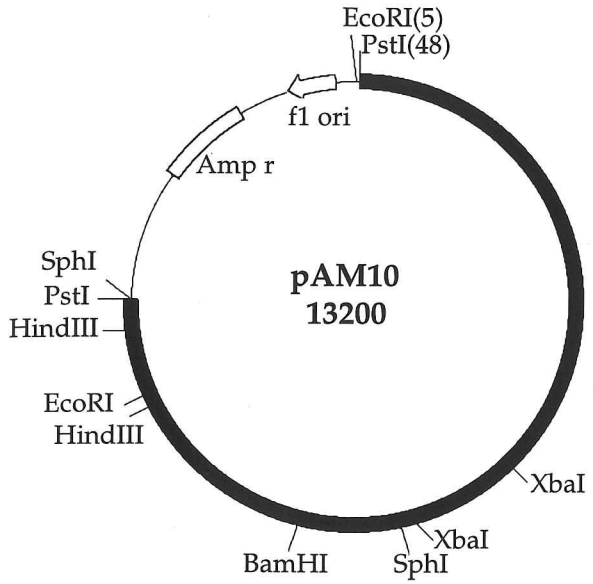


Figure 5.7 Restriction map of plasmid construct pAM10.
The 10 kb fragment (black box) cloned into the *Pst*I site of the vector pGEM-3Zf(+). The ampicillin-resistance gene is represented as Amp r and the f1 origin and its orientation are also indicated. Position 5 marks the first site in the pGEM-3Zf(+) polylinker. *am1* (the filled box) runs anticlockwise in this construct, the first codon being within the *Eco*RI site indicated in the fragment.

Subclones were constructed from the gel-purified *Xba*I pAM15 fragments and pGEM-3Zf(+) vector as previously described. The 7.2 kb fragment already contained linearised vector DNA so it was simply recircularised in a ligation reaction to form a subclone containing 4 kb of 17A DNA (pAMsc1), 3.9 kb of which was duplicated in pAMscE. The other three fragments were cloned into the *Xba*I site of dephosphorylated pGEM-3Zf(+). Sequence data obtained from the ends of these three subclones using the T7 and Sp6 primers identified the 1 kb subclone (pAMsc2) as containing the remaining *am2* sequence data and positioned the two remaining *Xba*I fragments within pAM10. This subcloning strategy is shown in Figure 5.8.

Much of the DNA in construct pAMsc1 had already been sequenced as it is duplicated in pAMscE and pAM0.7. However, unidirectional deletions were made in pAMsc1 and the entire fragment sequenced to confirm the initial data. Initially the plasmid DNA used in the exonuclease III nested deletion reactions was purified from *E. coli* JM109 cells using the Perfect prep™ kit. This kit produces closed circular supercoiled plasmid DNA free from protein and contaminating RNA and small fragments of *E. coli* chromosomal DNA that would otherwise contribute to the total concentration of termini in the exonuclease III reaction. It also minimises the generation of nicked circular molecules which would be degraded by exonuclease III from the site of the nick. Exonuclease III deletions were carried out by treating the linearised plasmid with exonuclease III for different lengths of time according to the Nested Deletions Kit (Pharmacia) protocol. Blunt-ended molecules were generated by incubation with S1 nuclease and the reaction terminated on addition of the chelating agent EDTA. Reactions were carried out until approximately 3.5 kb of insert DNA was deleted. Half of the reaction mixture for each time point was then subjected to agarose gel electrophoresis (0.5% agarose, 1x TBE), to determine the extent of the deletions,

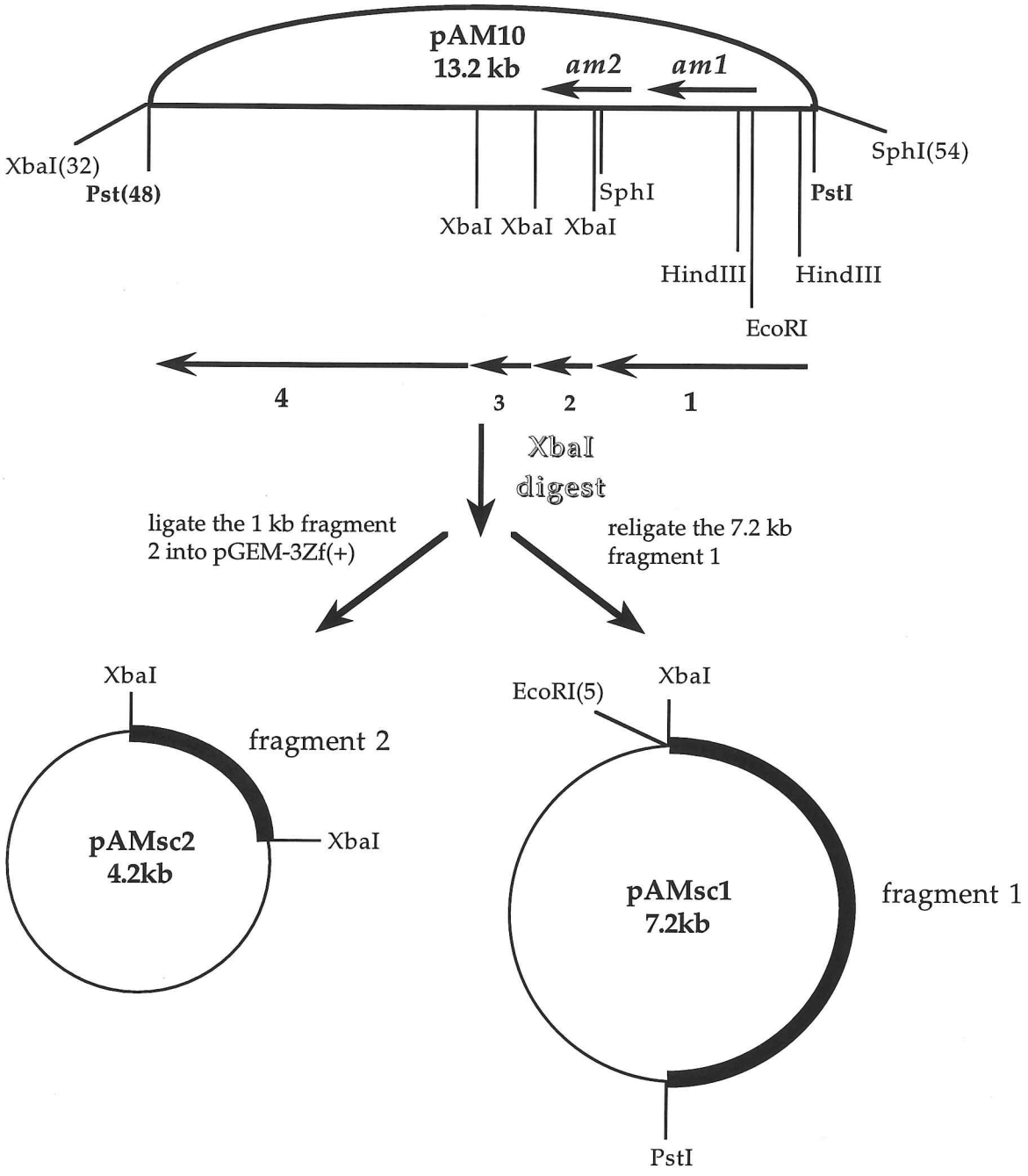


Figure 5.8 Subcloning of four *Xba*I fragments from pAM10.
Fragment 1 consists of 4.0 kb derived from pAM10 (shown as arrow 1) attached to 3.2 kb derived from the pGEM-3Zf(+) vector used to construct pAM10. Fragments 2, 3 and 4 (indicated by arrows) comprise 1, 1.2 and 4.2 kb fragments, repectively, derived from pAM10 and used to construct subclones pAMsc2, pAMsc3 and pAMsc4. For clarity only pAMsc2 is shown in this Figure.
The position and orientation of *am1* and *am2* are indicated on pAM10. Numbers in brackets after the restriction sites indicate their positions in the original vector before addition of the cloned fragment.

while the other half was recircularised by ligation and transformation into electrocompetent *E. coli* JM109.

In theory, transformants should have been generated carrying truncated fragments of different sizes which could be selected for by ampicillin resistance. The nested deletions and sequencing strategy is illustrated in Figure 5.9. However, although initial attempts with this protocol were successful in generating deletions of the correct size, very few if any transformants were obtained for each timepoint. It was thought that the problem lay in the concentration of plasmid DNA used in the initial reaction. The protocol recommends that at least 2.5 µg of DNA is used so that a final DNA concentration of 0.1 µg/µl is obtained. The miniprep plasmid DNA concentration used in this reaction was determined by its absorbance at 260 nm using a Pharmacia Biotech GeneQuant apparatus and exactly the recommended concentration used. However, the protocol also states that in the electrophoretic analysis of the deletion reaction the bands should be clearly visible; this had not been the case as the bands could barely be seen on the gel. It was therefore decided that a likely reason for failure of the transformation step was that there was insufficient DNA for successful ligations. Since no precipitation step where DNA could have been lost was involved it was determined that the initial concentration would have to be increased to above that stated in the protocol.

However, using the current method of DNA extraction the concentration of plasmid DNA eluted from the miniprep columns was only about 0.1 µg/µl. This falls far short of the maximum yield of the column of 0.6 µg/µl stated in the Perfect prep™ protocol. This low yield was thought to have been due to the choice of *E. coli* strain. *E. coli* JM109 contains large amounts of carbohydrate which can interfere with the extraction of plasmid DNA and reduce yields (Ausubel, 1991). Construct pAMsc1 DNA was therefore retransformed into *E. coli* TG2 and plasmid DNA extracted from this strain, again using the Perfect prep™ miniprep kit. Yield from the Perfect prep™ kit was increased to about 0.3-0.5 µg/µl using this strain but when the deletion reactions were repeated using this as the starting concentration the number of successful transformants was not significantly improved. Eventually, the method of plasmid DNA preparation was changed to use Qiagen Tip 20s (Section 2.2.1c) which also produces supercoiled plasmid DNA suitable for deletion reactions. Using this method a similar final yield of plasmid was obtained but as this method includes a precipitation step the DNA could be resuspended in a much smaller volume (20 µl) than that needed to elute the DNA from a Perfect prep column (50 µl) so concentrations of 1.0 µg/µl and above were obtained.

Using a starting concentration of about 1 µg/µl (ten times that recommended in the protocol) deletions were made in the pAMsc1 using the enzymes *Kpn*I (nuclease resistant) and *Xba*I (nuclease sensitive) to cut the appropriate sites in the vector polylinker. At this concentration, DNA from each time point was clearly visible on the analytical gel,

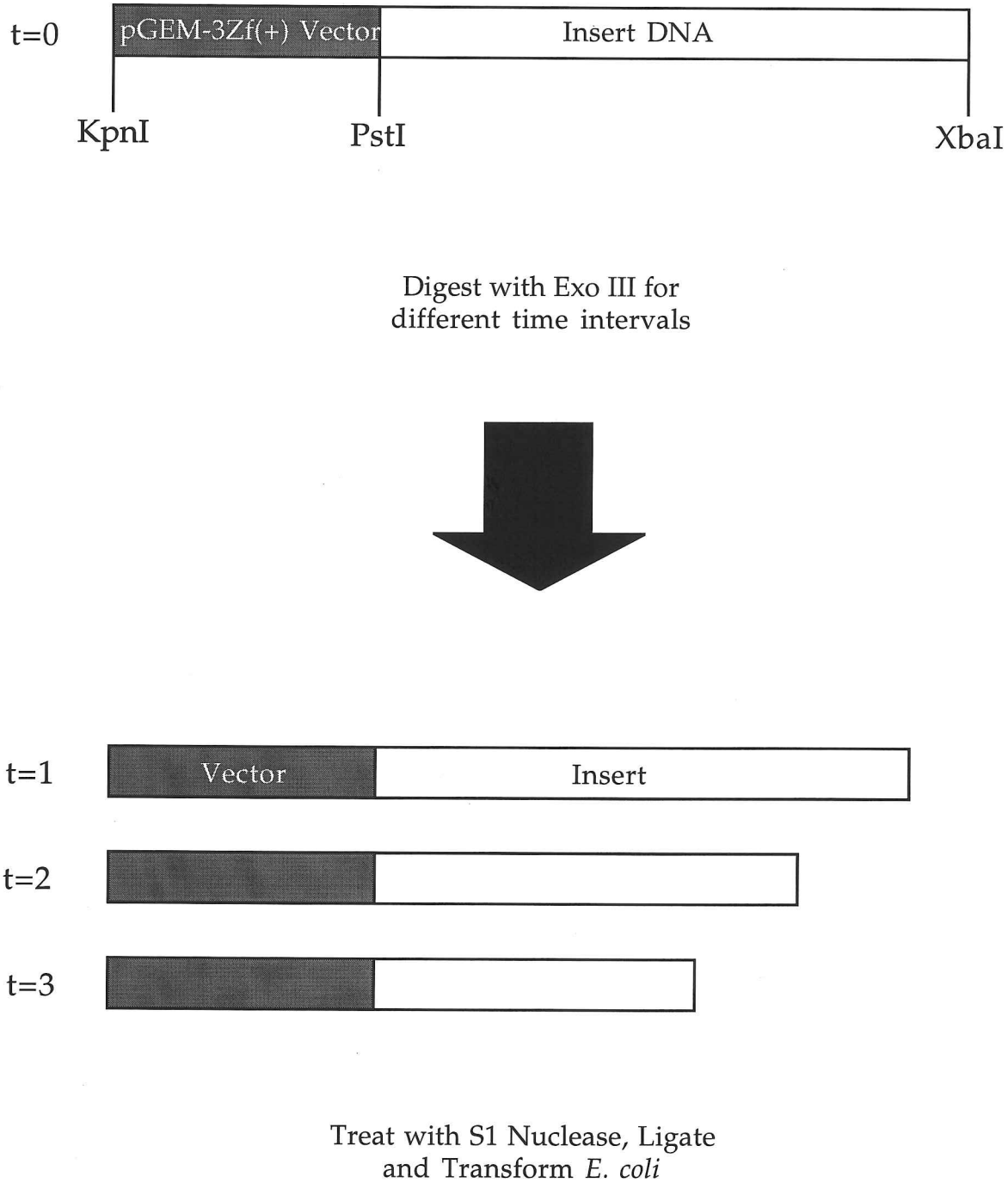


Figure 5.9 Strategy used to determine the nucleotide sequence of subclone pAMsc1.

transformants were obtained and DNA from each time point was prepared and sent for automated sequencing using the T7 primer.

Due to its small size pAMsc2 was sequenced by "gene walking" as previously described. In order to sequence across the breakpoint between subclones pAMsc1 and pAMsc2, oligonucleotides were designed to pAMsc1 and pAMsc2 sequence approaching the breakpoint in both directions. These oligonucleotides were then used as primers in sequencing reactions on the clone pAM10. The resulting sequences were compiled using the 'bestfit' program from GCG. Along with the sequence obtained from pAMscE and pAM0.7, this generated 4.8 kb of unambiguous sequence containing the complete coding regions for the putative crystal proteins AM1 and AM2, potential upstream control regions and downstream terminator structures and the gap non-coding region between the *am1* and *am2*. The N-terminal sequence (YPLANDPQA) from the trypsin activated protein that was obtained in Chapter 3 can be seen in this sequence from residues 29 to 37 of AM1. The start of this sequence is marked in the text. The compiled sequence is shown in Figure 5.10.

From this information the *am2* gene was found to encode a 542 amino acid residue protein of predicted mass 61,839 Da. A search for a downstream terminator structure revealed the possibility of a potential hairpin loop (4637-4656). The amino acid sequence of AM2 was aligned with the C-terminal regions of Cry4A and Cry4B using the Clustal method with PAM250 residue weight table. The amino acid sequence of the Cry10A ORF2 was also included for comparison, although only the first 238 residues are known. Unlike the N-terminal alignment in Figure 5.4 it was not necessary to edit this C-terminal alignment as all obvious regions of sequence conservation had been correctly aligned. The percentage similarities for the AM2 alignment (Figure 5.11) are shown in Figure 5.12. AM2 shares over 60% similarity with the C-terminal regions of Cry4A and Cry4B and 55% to the Cry10A ORF2. The obvious similarity between these four regions of sequence indicates that the two-gene configurations of AM1 and Cry10A have evolved from genes that would otherwise encode Cry4A/B proteins of about 130-135-kDa. A BLAST search was carried out on the AM1/AM2 intergenic region to explore the possibility that the disruption could have been caused by an Insertion Sequence but no similarity to any previously identified Insertion Sequences was found.

It is interesting to note that the Cry4A and Cry4B C-terminal regions are very highly conserved, showing over 90% similarity. This observation is obviously not new and the high degree of sequence conservation between the C-terminal regions of Cry4A and Cry4B has always indicated that this region must play an important functional role. However, the importance of this region is highlighted when the new sequence information from AM1 and that available from Cry10A is considered. It also raises interesting phylogenetic considerations.

```
TGC AGA ATT AAA TAT TCA AGC TAA TTA TAA TAA TGT TCA AGT AAG ATT AAC AAG ACT AGG 60
PstI(1)
CCC TTT AAC ATA GAT AAG TGT GTT GGT TCT GTT GCA AAC CTT TAA ATT CTC TAC ACT TAG 120

CCT TTA AAA AAA GAA ACA ACC TCT ACT ATA ATT GAA GTT GGA TTT GCT TAA CTT GAT AGC 180

GAT GTG GTG ATA TCC CTA ATT ATG TTG GAT ATA TTA AAG TTT AGT AAT TGC ATT ATG AAT 240

TAA AAC TAT TGA GTA AAA AAT CCG ACT CAT AAG CTT TAA GTC GGA TTT TTA TTT ATA ATT 300
                                    HindIII(272)
TTT TAT AAA GTT TCC CTA AAT CTT TCT TTA TCT GAG ATT TCA GTT GAT GAT TTT CGT CTT 360

CTA GTG TCT TAA CTT TCT GTC TTA GTA GCT TCA TTT TAT CTT TTA AAG TTT CTA TAA TTA 420

CAG TTT TCT GTA CAC AGG TAA AAT TCC CAT AAT TCG GTT TTT CCT AGA AAA AAT AAA ATG 480

AAT CCA ATA CAG GTT CGG ATT TAT TTT CTA CAA TTT CTT TCA ATG ACG AAT AAT TAA ATA 540

TCT ACA ACA CTC TAT GTA GTA TCC CTC TCT TTT TTA GAG GGG GGG AAA AAA CAT CCA AGG 600

GTG AAT TTC GTT CAT ATA AAA GTG AAT GAC TCT TTT CGC ACC CTT AAA AAC AAC AAA GAA 660

AAA ATC GTT CTA TAG AAA TCT GAA ACT TCA AAA AAT TAT ATG CAA TAC ATA AAG AAA AGG 720

TTT AAA AAA TCA ATA TTT TTA CCA AAA ATA ATG GGT TTA TTT GTA GAA ACA TTG TCA CAG 780

GAA TAC ATT GGG GCA CTA CGA ATA TAT AAC AAG ACA CCT AGC ATA TAC TGT TTG GGT ATC 840
                          -35                               -10
                                                  M   N   S   Y   Q   N   K   N   E   9
TAA AAA TAA GGA CCA CAT AAG GGA GTG AAA AAT ATG AAT TCA TAC CAA AAT AAA AAT GAA 900
                          RBS                   EcoRI(877)
                                                                              ######
Y   E   I   L   D   A   S   R   N   N   S   N   M   S   N   R   Y   P   R   Y   29
TAT GAA ATA TTG GAT GCT TCA CGA AAC AAC TCT AAT ATG TCT AAC CGT TAT CCA AGG TAT 960

P   L   A   N   D   P   Q   A   C   M   Q   N   T   N   Y   K   D   W   L   A   49
CCA CTA GCA AAT GAT CCA CAA GCT TGT ATG CAG AAT ACG AAT TAT AAA GAT TGG TTG GCT 1020
                      HindIII(981)
T   C   N   G   T   T   V   P   F   S   N   P   S   Q   L   L   K   V   G   G   69
ACG TGC AAC GGA ACG ACC GTT CCT TTT TCT AAT CCT TCA CAA CTA CTT AAA GTC GGA GGA 1080

I   V   V   Q   R   V   L   G   G   I   A   T   F   F   P   G   I   G   P   L   89
ATT GTA GTT CAG AGG GTT TTG GGA GGT ATC GCT ACC TTT TTT CCT GGG ATT GGT CCT CTT 1140

V   P   F   L   T   F   F   A   S   L   L   W   P   S   G   S   S   G   N   D   109
GTG CCC TTT TTA ACT TTT TTC GCT AGT TTG CTT TGG CCT AGC GGA TCA TCA GGG AAT GAT 1200

I   W   E   K   L   M   K   E   V   A   D   L   I   Q   Q   E   L   T   T   Y   129
ATT TGG GAA AAG TTG ATG AAA GAA GTT GCG GAT TTG ATA CAG CAA GAA TTA ACT ACT TAT 1260
```

Figure 5.10 Nucleotide sequence of 4753 bp of Bt ssp. *fukuokaensis* 17A DNA encoding the 65-kDa crystal protein gene (AM1) and the second open reading frame (AM2).
The 2097 bp open reading frame *am1* is shown from base 873 to base 2970. The 1608 bp open reading frame *am2* is shown from base 3040 to base 4648. Putative Bt promoter sequences are indicated as well as a ribosome binding site (RBS). Selected restriction sites are shown alongside their position in the nucleotide sequence. On this page the *Pst*I site marks the start of the sequence and the site derived from pAM10 in the subclone pAMsc1 (see Figure 5.9). The two *Hin*dIII sites mark the limits of the fragment cloned in pAM0.7 and the *Eco*RI site is the start of the fragment cloned in pAMscE. The start of the N-terminal sequence of the trypsin activated protein is marked with hatching (######).

```
T   I   D   K   A   T   A   E   L   G   G   L   K   E   L   L   D   S   Y   N   149
ACA ATA GAT AAA GCA ACC GCT GAG TTA GGC GGA TTA AAA GAA CTA TTG GAT TCA TAT AAC 1320

R   A   F   A   S   W   E   V   G   N   A   T   P   G   L   V   K   G   Y   I   169
AGA GCT TTT GCT TCA TGG GAA GTA GGC AAT GCA ACT CCT GGG CTG GTA AAA GGT TAT ATA 1380

E   S   L   H   R   K   F   V   D   K   V   I   S   G   F   T   Q   P   G   Y   189
GAG AGC CTT CAT CGT AAG TTT GTG GAT AAG GTT ATA AGC GGT TTC ACG CAA CCA GGT TAT 1440

E   K   I   L   L   P   S   Y   A   I   A   A   N   F   H   L   L   V   L   R   209
GAA AAA ATA TTA TTA CCT TCA TAT GCA ATT GCC GCG AAT TTT CAT TTG CTG GTA TTA CGT 1500

D   I   E   I   Y   G   E   R   L   G   F   S   Q   V   D   R   N   F   Y   N   229
GAC ATT GAA ATT TAT GGA GAA AGA CTA GGT TTT TCT CAA GTG GAC CGA AAT TTT TAT AAT 1560

C   E   L   K   F   F   M   A   K   Y   T   N   Y   C   V   D   T   Y   N   K   249
TGT GAA CTA AAG TTT TTT ATG GCA AAG TAT ACG AAT TAT TGC GTA GAT ACA TAC AAT AAA 1620

G   L   A   S   E   K   E   K   G   W   V   P   F   H   R   Y   R   R   E   M   269
GGT TTG GCT TCA GAA AAA GAA AAG GGT TGG GTG CCT TTC CAT CGA TAT CGT AGA GAA ATG 1680

T   L   A   V   L   D   I   I   A   L   F   P   L   Y   D   A   R   L   Y   P   289
ACT TTA GCT GTA TTA GAT ATA ATT GCA TTA TTC CCA CTC TAT GAT GCA AGA CTC TAT CCA 1740

A   T   D   N   K   E   M   P   V   K   S   E   L   T   R   E   I   Y   S   D   309
GCT ACG GAT AAT AAA GAA ATG CCA GTT AAA TCC GAA TTG ACT CGG GAA ATT TAT TCG GAT 1800

V   I   N   S   D   R   F   G   I   V   P   P   Y   N   Y   S   Q   N   E   E   329
GTC ATT AAT AGC GAT AGG TTC GGA ATT GTA CCC CCT TAT AAT TAT TCT CAA AAC GAA GAA 1860

R   Y   T   R   P   P   H   L   F   T   W   L   R   G   L   D   F   V   T   N   349
CGT TAT ACA CGA CCA CCT CAT CTC TTC ACT TGG TTA CGA GGG CTT GAC TTT GTA ACC AAT 1920

V   L   T   S   G   T   Y   A   Y   R   W   R   V   L   T   G   C   Q   N   K   369
GTT TTG ACT AGC GGA ACT TAT GCT TAT AGA TGG CGT GTT TTA ACT GGT TGT CAA AAT AAA 1980

Y   S   Y   T   R   G   N   G   T   I   T   G   P   F   R   G   Y   P   V   E   389
TAT TCT TAT ACC AGA GGC AAT GGT ACT ATA ACT GGT CCT TTT CGG GGT TAT CCT GTA GAG 20409

S   G   G   S   T   S   N   I   T   I   E   E   G   S   Y   I   Y   N   L   L   406
TCT GGT GGA AGC ACT TCT AAC ATT ACT ATT GAA GAA GGC TCC TAC ATT TAT AAC TTG TTG 2100

P   R   S   F   E   Y   I   S   P   W   Y   F   T   T   N   I   A   M   I   T   429
CCA AGA AGC TTT GAA TAT ATT TCC CCT TGG TAT TTT ACG ACA AAT ATT GCA ATG ATT ACT 2160

F   L   L   T   N   N   N   S   S   T   E   K   V   Y   G   S   L   V   N   Q   449
TTC TTG CTA ACA AAT AAT AAT AGT TCA ACA GAA AAA GTT TAT GGT TCT CTA GTA AAC CAA 2220

P   N   L   P   I   V   Q   T   D   F   D   F   L   T   N   K   E   V   T   G   469
CCT AAT TTA CCT ATT GTT CAA ACA GAT TTT GAT TTT CTT ACA AAT AAA GAA GTA ACT GGA 2280

S   P   T   Y   N   N   Y   N   H   I   L   S   Y   M   L   L   G   Y   D   W   489
TCT CCA ACA TAC AAT AAC TAT AAT CAT ATT TTG TCA TAC ATG TTG CTA GGT TAT GAT TGG 2340

N   Q   T   G   G   I   G   T   H   G   Y   S   F   A   F   T   H   S   S   V   509
AAC CAG ACG GGT GGA ATA GGC ACA CAT GGA TAT TCA TTT GCA TTT ACG CAT AGT AGC GTT 2400

D   P   Y   N   T   I   A   S   D   K   I   T   Q   I   P   A   V   K   G   H   529
GAT CCT TAT AAC ACC ATT GCC TCA GAT AAA ATT ACA CAA ATA CCA GCT GTG AAA GGG CAT 2460
```

Figure 5.10 continued.

```
S   L   Q   N   G   I   V   V   G   G   P   G   H   T   G   G   D   L   V   N   549
TCT CTT CAA AAT GGA ATC GTT GTA GGA GGT CCC GGT CAT ACA GGT GGA GAT TTG GTT AAT 2520

M   G   Y   A   S   T   L   S   M   N   C   Y   F   S   Q   P   L   N   Y   K   569
ATG GGG TAT GCG AGT ACG CTC TCG ATG AAC TGT TAT TTT TCT CAA CCA CTA AAT TAT AAA 2580

M   R   I   R   Y   S   T   S   F   Y   T   P   F   Y   I   S   S   P   H   R   589
ATG CGT ATT CGT TAT TCG ACA AGT TTT TAT ACT CCA TTT TAT ATA TCT TCA CCA CAT CGT 2640

S   G   V   V   S   I   S   L   F   Q   T   N   Y   V   K   V   D   G   K   N   609
TCA GGG GTG GTA TCA ATA TCT CTA TTC CAA ACA AAC TAT GTA AAG GTT GAT GGT AAA AAT 2700

I   E   V   D   L   M   D   P   R   A   F   Q   I   I   E   V   P   V   E   F   629
ATT GAG GTG GAC CTG ATG GAT CCT CGT GCA TTC CAG ATT ATA GAA GTT CCT GTG GAG TTT 2760
                    BamHI(2719)
Q   A   T   S   S   G   Y   A   N   L   I   F   T   A   N   A   S   Y   G   V   649
CAG GCA ACT TCT TCA GGA TAT GCA AAT CTT ATT TTT ACG GCT AAT GCT TCT TAT GGT GTG 2820

Y   I   D   K   I   E   F   I   P   S   N   I   Q   I   D   K   C   T   K   C   669
TAT ATT GAT AAA ATT GAA TTT ATC CCA AGT AAT ATC CAA ATT GAT AAA TGT ACG AAA TGT 2880

Q   F   E   E   K   V   C   T   C   R   C   E   G   V   Q   S   L   E   T   E   689
CAA TTC GAA GAA AAA GTA TGT ACA TGT AGA TGT GAA GGA GTA CAA TCC TTA GAA ACA GAA 2940

K   E   I   V   N   S   L   F   I   N   *                                       699
AAA GAG ATT GTA AAT AGT TTA TTT ATC AAT TAA AAC AAA GTA GGT ACT GAC GTA GGA GGT 3000
                                                                        RBS
                                                    M   Y   T   N   T   M   K   7
AAG GCT GTT CGA AAA ATA AGT AGA AAA GGT AGT GAA TAC TAT GTA TAC CAA TAC GAT GAA 3060

 N   K   L   K   I   E   T   T   D   Y   E   I   D   Q   A   A   I   S   I   E   37
AAA TAA ATT AAA AAT AGA AAC GAC AGA TTA TGA AAT AGA TCA AGC GGC TAT TTC TAT AGA 3120

 C   M   S   N   E   H   S   P   K   E   K   M   I   L   W   D   E   V   K   Q   47
ATG TAT GTC TAA TGA ACA TTC TCC CAA AGA AAA AAT GAT ATT ATG GGA TGA AGT AAA ACA 3180

 A   K   Q   L   S   W   S   R   N   L   L   Y   N   G   D   F   E   D   A   S   67
GGC AAA ACA ACT TAG CTG GTC TCG TAA TTT ACT CTA CAA TGG TGA TTT TGA AGA TGC ATC 3240

 S   G   W   K   T   S   N   T   I   E   I   R   E   D   S   P   I   F   K   G   87
AAG CGG CTG GAA AAC AAG TAA TAC GAT TGA GAT TCG AGA GGA TAG TCC CAT TTT TAA AGG 3300

 H   Y   L   H   M   F   G   A   R   D   I   A   G   T   L   F   P   T   Y   L   107
ACA TTA TCT TCA TAT GTT TGG GGC AAG AGA TAT TGC TGG AAC CCT ATT TCC AAC CTA TCT 3360

Y   Q   K   I   E   E   S   K   L   K   P   Y   T   R   Y   R   V   R   G   F   127
GTA TCA AAA AAT AGA GGA ATC CAA ATT AAA ACC CTA TAC ACG TTA TCG AGT AAG AGG ATT 3420

 V   E   S   S   K   D   L   K   L   V   V   T   R   Y   G   K   E   I   D   A   147
TGT GGA AGG TAG TAA AGA TCT AAA ATT AGT GGT AAC ACG TTA CGG GAA AGA AAT TGA CGC 3480

 I   M   N   V   P   N   D   L   S   Y   M   Q   P   N   P   S   C   G   D   Y   167
CAT TAT GAA TGT TCC AAA TGA TTT GTC CTA TAT GCA GCC TAA TCC TTC ATG TGG AGA TTA 3540

 R   C   E   S   S   S   Q   Y   V   S   Q   G   Y   L   T   P   V   T   D   G   187
TCG CTG TGA ATC ATC GTC CCA GTA TGT CAG CCA AGG TTA TCT TAC GCC AGT AAC AGA TGG 3600

 Y   A   S   D   R   Y   A   C   Q   S   D   R   G   K   K   H   V   K   C   H   207
ATA TGC TTC TGA TAG GTA TGC ATG CCA GTC CGA TCG AGG TAA AAA GCA TGT GAA GTG TCA 3660
                        SphI(3621)
```

Figure 5.10 continued. The end of *am1* and the start of *am2*. * indicates a stop codon. A ribosome binding site (RBS) is indicated. The *Sph*I site marks the end of the fragment in pAMscE.

```
 D   R   H   P   F   D   F   H   I   D   T   G   E   V   D   A   N   T   N   I  227
CGA TCG TCA TCC ATT TGA TTT TCA TAT TGA CAC CGG AGA AGT AGA TGC AAA TAC AAA CAT 3720

 G   I   D   V   L   F   K   I   S   N   P   D   G   Y   A   T   V   G   N   L  247
AGG TAT TGA TGT CTT GTT TAA AAT TTC TAA TCC AGA TGG ATA CGC TAC AGT AGG GAA TCT 3780
                                                                    XbaI(3779)
 E   V   I   E   E   G   P   L   T   G   E   A   L   A   H   V   K   Q   K   E  267
AGA AGT CAT TGA AGA AGG ACC ACT AAC AGG CGA AGC ATT AGC ACA TGT GAA ACA AAA GGA 3840

 K   K   W   K   Q   Y   M   E   K   K   R   W   E   T   Q   Q   A   Y   D   P  287
AAA GAA ATG GAA ACA ATA CAT GGA GAA AAA ACG TTG GGA AAC ACA ACA AGC TTA TGA TCC 3900

 A   K   Q   A   V   D   A   L   F   T   N   A   Q   G   E   E   L   H   Y   H  307
GGC AAA ACA AGC AGT AGA TGC ATT ATT CAC AAA TGC ACA AGG AGA AGA GTT ACA CTA TCA 3960

 I   T   L   D   H   I   Q   N   A   N   Q   L   V   Q   S   I   P   Y   V   H  327
TAT TAC TTT AGA TCA TAT TCA GAA CGC CAA TCA GTT GGT ACA GTC GAT TCC TTA TGT ACA 4020

 H   A   W   L   P   D   A   P   G   M   N   Y   D   L   Y   N   N   L   K   V  347
CCA TGC TTG GTT ACC GGA TGC TCC AGG TAT GAA CTA TGA TTT ATA TAA CAA TTT AAA GGT 4080

 R   I   E   Q   A   R   Y   F   Y   D   A   R   N   V   I   T   N   G   D   F  367
ACG TAT AGA ACA AGC ACG CTA TTT CTA TGA TGC ACG GAA TGT CAT AAC AAA TGG TGA CTT 4140

 T   Q   G   L   Q   E   W   H   A   T   G   K   A   A   V   Q   Q   M   D   G  387
TAC ACA AGG ATT ACA GGA ATG GCA CGC AAC AGG AAA GGC CGC GGT GCA ACA AAT GGA TGG 4200

 A   S   V   L   V   L   S   N   W   S   A   G   V   S   Q   N   L   Y   A   Q  407
CGC TTC TGT ATT AGT TCT ATC AAA CTG GAG TGC TGG GGT ATC TCA AAA TCT GTA TGC CCA 4260

 D   H   H   G   Y   V   L   R   V   I   A   K   K   E   G   P   G   K   G   Y  427
AGA TCA TCA TGG ATA TGT GTT ACG TGT GAT TGC CAA AAA AGA AGG TCC TGG AAA AGG ATA 4320

 V   T   M   M   D   C   N   G   H   Q   E   T   L   R   F   T   S   C   E   K  447
CGT AAC GAT GAT GGA TTG TAA TGG TCA TCA GGA AAC ACT TAG GTT TAC TTC TTG TGA AAA 4380

 G   Y   M   T   K   T   V   E   V   F   P   E   S   D   R   V   R   I   E   I  467
AGG ATA TAT GAC AAA AAC AGT AGA GGT ATT CCC AGA AAG TGA TCG TGT ACG GAT TGA AAT 4440

 G   E   T   E   G   T   F   Y   I   D   S   I   E   L   L   C   M   Q   G   Y  487
AGG AGA AAC CGA AGG TAC ATT TTA TAT AGA TAG CAT CGA GTT GCT TTG TAT GCA AGG ATA 4500

 A   S   N   N   N   P   H   T   D   N   M   Y   E   Q   S   Y   N   G   N   Y  507
TGC TAG CAA TAA TAA CCC GCA CAC GGA TAA TAT GTA TGA GCA AAG TTA TAA TGG AAA TTA 4560

 T   H   N   D   D   L   H   S   G   C   T   C   N   Q   G   H   N   S   G   C  527
TAC TCA CAA TGA TGA CCT GCA TTC CGG TTG CAC ATG TAA CCA AGG GCA TAA CTC TGG CTG 4620

 T   C   N   Q   G   Y   N   R   *                                              537
TAC ATG TAA TCA AGG ATA TAA CCG TTA ACG ATT CTA AAT AAG AAT CAA CAT CAT TGC GAA 4680
                          ----------><----------
ATA TAA AAA CCT ACT CAC AAA ATC TAT TGC ATA TCA TAA CAT AAG CTT TAC AAA TAA CTG 4740

ACA TAT TCT AGA
      XbaI(4748)
```

Figure 5.10 continued.
*Xba*I(3779) marks the end of the DNA fragment in pAMsc1 and the start of that in pAMsc2, the end of which is marked by *Xba*I(4748). * indicates a stop codon. A predicted hairpin loop structure is also shown by opposing arrows.

```
AM2                                                    VEKVVNTMYTNTMKNKLKIETTDYEID
Cry4A      DVYTNTTVLIDKIEFLPITRSIREDREKQKLETVQQIINTFYANPIKNTLQSELTDYDID
Cry4B      NMTSNNQVIIDRIEIIPITQSVLDETENQNLESEREVVNALFTNDAKDALNIGTTDYDID
Cry10A                                                 ARKVVNPMFTSGAKNRLKLETTDYEID
                                                        ***    **  ** *  ********

AM2        QAAISIECMSNEHSPKEKMILWDEVKQAKQLSWSRNLLYNGDFEDASSGWKTSNTIE
Cry4A      QAANLVECISEELYPKEKMLLLDEVKNAKQLSQSRNVLQNGDFESATLGWTTSDNIT
Cry4B      QAANLVECISEGIISKEKMLLLDEVKNAKQLSQSRNVLQNGDFESRTLGWTTSDNIT
Cry10A     QVANAIECMSDEQYSKEKLMLWDQVKHAKYLSQSRNLLQNGDFEDVFHGWTTSDHMY
           ***  *** * *   **** * **** ***** *** * *****    ** *** *

AM2        IREDSPIFKGHYLHMFGARDIAGTLFPTYLYQKIEESKLKPYTRYRVRGFVESSK
Cry4A      IQEDDPIFKGHYLHMSGARDIDGTIFPTYIFQKIDESKLKPYTRYLVRGFVGSSK
Cry4B      IQEDDPIFKGHYLHMSGARDIDGTIFPTYIFQKIDESKLKPYTRYLVRGFVGSSK
Cry10A     IQSDNSTFKGNYLNISGARDIYLTIFPTYIYQKIDESKLKPYTRYLVRGFVGSSK
           * ** ********** ***** ** ****  ************** ***** ***

AM2        DLKLVVTRYGKEIDAIMNVPNDLSYMQPNPSCGDYRCESSSQYVSQGYLTPVTDGY--A
CryIVA     DVELVVSRYGEEIDAIMNVPADLNYLYPST----FDCEGSNRCETSA--VPANIGN--T
CryIVB     DVELVVSRYGEEIDAIMNVPADLNYLYPST----FDCEGLIVVSVRCA-AN----IWDT
CryIVC     DVELVVSRYGKEIDTVMNVPFDIPYV--SS----RPCNELYDGEQQPY-PNGNVGYYNP
           *  ******* ********* ** *  **       **                 *

AM2        SDR-YACQSDRG-------KKHVKCHDRHPFDFHIDTGEVDANTNIGIDVLFKISNPD
Cry4A      SDML YSCQYDTG-------KKHVVCQDSHQFSFTIDTGALDTNENIGVWVMFKISSPD
Cry4B      SDML YSCQYDTG-------KKHVVCQDSHQFSFTIDTGALDTNENIGVWVMFKISSPD
Cry10A     MSAFYPS-YTSDARQCMPRKKQIVCQDFHQFKFHIDTGEVDYNTNIGIWVMFKIS
           **  * ** * *        **** * * * * * ****  * * **** * *******

AM2        GYATVGNLEVIEEGPLTGEALAHVKQKEKKWKQYMEKKRWETQQAYDPAKQAVDA
Cry4A      GYASLDNLEVIEEGPIDGEALSRVKHMEKKWNDQMEAKRSETQQAYDVAKQAIDA
Cry4B      GYASLDNLEVIERGPIDGEALSRVKHMEKKWNDQMEAKRSETQQAYDVAKQAIDA
           ****  *********  ***** **  ****   ** ** ******* *******

AM2        LFTNAQGEELHTHITLDHIQNANQLVQSIPYYHHAWLPDAPGMNYDLYNNLKVRIE
Cry4A      LFTNVQDEALQFDTTLAQIQYAEYLVQSIPYYYNDWLSDVPGMNYDIYVELDARVA
Cry4B      LFTNVQDEALQFDTTLAQIQYAEYLVQSIPYYYNDWLSDVPGMNYDIYVELDARVA
           **** * * *    **  ** *  ********   ** * ****** *  *  **

AM2        QARYFYDARNVITNGDFTQGLQEWHATGKAAVQQMDGASVLVLSNWSAGVSQNLY
Cry4A      QARYFYDIRNIIKNGDFTQGVGMWHVTGNADVQQIDGVSVLVLSNWSAGVSQNVH
Cry4B      QARYFYDTRNIIKNVDFTQGVGMWHVTGNADVQQIDGVSVLVLSNWSAGVSQNVH
           **** ** **** *******   ** ** * *** ** ***************
```

Figure 5.11 Clustal alignment of the amino acid sequence of AM2 with the C-terminal regions of Cry4A and Cry4B (continued overpage).
The Cry10A ORF2 is also included as far as it has been sequenced. The alignment was generated using a PAM250 residue weight table. Conserved or semiconserved residues are shown in bold and marked with an asterisk. Residues that are conserved between AM2 and all but one other sequence are marked with an asterisk in plain text.

```
AM2       AQDHHGYVLRVIAKKEGPGKGYVTMMDCNGHQETLRFTSCEKGYMTKTVEVFPES
Cry4A     LQHNHGYVLGVIAKKEGPGNGYVTLMDWEENQEKLTFTSCEEGYITKTVDVFPDT
Cry4B     LQHNHGYVLRVIAKKEGPGNGYVTLMDCEENQEKLTFTSCEEGYITKTVDVFPDT
           *  ***********  ****  ***   ** * ***** ** **********

AM2       DRVRIEIGETEGTFYIDSIELLCMQGYASNNNPHTDNMYYEQSYNGNYTHNDDLHSGC
Cry4A     DRVRIEIGETEGSFYIESIELICMNE
Cry4B     DRVRIEIGETEGSFYIESIELICMNE
          ********************* **

AM2       TSGCTCNQGYNRCNQGHN
Cry4A
Cry4B
```

Figure 5.11 continued.

5.4 Phylogenetic Analysis of AM1 in Relation to Cry4A, Cry4B and Cry10A

For the basis of this analysis the two open reading frames of AM1 and Cry10A will be considered to be the N and C-terminal regions of full-length Cry4A/B type 130-kDa δ-endotoxins. First, if just the N-terminal regions of the four proteins (Cry4A, Cry4B, Cry10A and AM1) are considered, a Clustal alignment (manually edited as before) indicates that all four proteins show very similar levels of sequence conservation (Figure 5.12a). The phylogenetic tree in Figure 5.13a shows more clearly that based on this N-terminal region, the proteins can be considered to have diverged from a common ancestor at a very similar point in evolutionary history.

However, when the C-terminal regions are aligned (Figure 5.12b and 5.13b) Cry4A and Cry4B share a much greater similarity and appear to have diverged more recently. This discrepancy can be explained if it is considered that unconstrained by the need to retain functional properties, the disrupted C-terminal regions of Cry10A and AM1 have been subject to more rapid evolutionary change than the functional C-terminal regions of 4A and 4B. The almost total C-terminal sequence conservation of Cry4A and 4B has long been an indicator that whatever the role of this region it is strongly sequence dependent. The comparable lack of conservation in the redundant C-terminal regions of AM1 and Cry10A further supports this theory.

It is, however, perhaps hasty to dismiss the C-terminal regions of AM1 and Cry10A as redundant and non-functional. Despite the sequence diversion that seems to have taken place, it appears that *am2* has retained its integrity as a complete open reading frame. Although the full sequence of *cry10A ORF2* has never been determined the detection of a 60-kDa protein in Bti inclusions with an identical N-terminus to that of the *ORF2* (Pucell, 1997) strongly suggests that this is also the case for this gene. The δ-endotoxin C-terminal

Percentage similarity

	AM1	Cry4A	Cry4B	Cry10A	
AM1		31.1	27.9	28.7	AM1
Cry4A	64.1		27.2	33.6	Cry4A
Cry4B	66.2	67.6		25.5	Cry4B
Cry10A	68.8	63.8	67.5		Cry10A
	AM1	Cry4A	Cry4B	Cry10A	

Percentage divergence

Figure 5.12a Protein sequence relationships between AM1, Cry10A and the corresponding N-terminal regions of Cry4A and 4B.

Percentage similarity

	AM2	Cry4A	Cry4B	Cry10A	
AM2		65.0	64.3	54.9	AM2
Cry4A	34.0		91.6	54.9	Cry4A
Cry4B	34.7	6.4		56.9	Cry4B
Cry10A	40.1	40.3	37.9		Cry10A
	AM2	Cry4A	Cry4B	Cry10A	

Percentage divergence

Figure 5.12b Protein sequence relationships between AM2, ORF2 of Cry10A and the corresponding C-terminal regions of Cry4A and 4B.

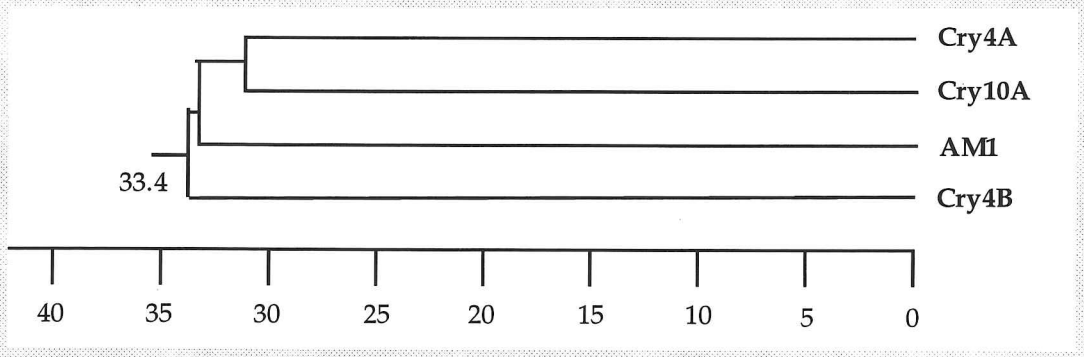


Figure 5.13a Phylogenetic tree based on the sequence relationships of the N-terminal regions described in Figure 5.12a.

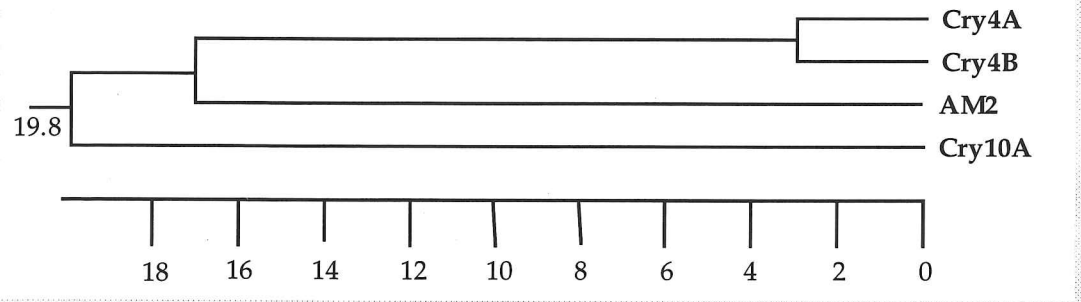


Figure 5.13b Phylogenetic tree based on the sequence relationships of the C-terminal regions described in Figure 5.12b.

The length of each pair of branches represents the distance between sequence pairs. The scale beneath the tree measures the distance between sequences; units indicate the number of substitition events.

region is thought to play an important role in stabilising and packaging the polypeptide chains into the crystal. Despite being expressed independently, the Cry10A C-terminal region may still be able to perform these functions and therefore play a role in the stability of the Bti inclusion as a whole. A ribosome binding sequence is also evident upstream of *am2*, indicating that this C-terminal region may be similarly expressed and that AM2 may play some role in the 17A native inclusion.

This theory is supported by the fact that the disruption between the two open reading frames of the *am1* and *cry10A* sequences occurs at almost exactly the same point in each gene. This implies that the disruption occurred to a common progenitor before sequence diversion gave rise to two different proteins. Should this be the case then it would be surprising that the C-terminal regions of these two proteins have remained more highly conserved than their N-terminal regions if they did not serve some functional role. However, it is also possible that the disruptions to AM1 and Cry10A occurred as independent events in two already divergent proteins, indicating that this particular region is a "hot spot" for mutation. This second theory would more easily explain the C-terminal region sequence conservation between these two proteins in comparison with the near identity of the same regions of Cry4A and 4B.

Based on the results of the independently aligned N and C-terminals (Figure 5.12) it was possible to "reconstruct" AM1 and Cry10A theoretically to form continuous open reading frames that align with Cry4A and 4B along their entire length. The reconstruction involves the removal of intervening regions of DNA (123 bp for AM1 and 83 bp for Cry10A), indicating that the disruption occurred by the insertion of a DNA fragment rather than by point mutations to create non-sense codons. The process of reconstruction is represented in Figure 5.15. The reconstructed AM1 (hereafter refered to as AM1/2) was then aligned with Cry4A and Cry4B along their entire length by Clustal, and manually edited to obtain the percentage similarity values shown in Figure 5.14. The reconstructed Cry10A (Cry10A/2) was also included in this alignment, although the lack of part of the most highly conserved region gives misleadingly low percentage similarity values.

As expected, this alignment shows that even reconstructed, AM1/2 is less closely related to Cry4A or 4B than they are to each other. It is, however, difficult to determine the degree of relation between these sequences on the basis of this information alone. In order to put these percentage similarity scores into phylogenetic context, AM1/2 was further aligned with a range of Cry proteins from representatives of different classes of δ-endotoxins. The results of this alignment are shown in Figure 5.16a and b as a percentage similarity table and a phylogenetic tree. It can be clearly seen that AM1/2 is more closely related to Cry4A and 4B than to any other δ-endotoxin type. Due to its complexity no attempt was made to edit this Clustal alignment, hence the discrepancies between the values in Figure 5.14 and those in Figure 5.16a.

Percentage Similarity

	AM1/2	Cry4A	Cry4B	Cry10A/2	
AM1/2		43.4	44.2	36.4	AM1/2
Cry4A	52.6		54.9	38.7	Cry4A
Cry4B	51.4	39.4		36.5	Cry4B
Cry10A/2	60.4	57.8	59.0		Cry10A/2
	AM1/2	Cry4A	Cry4B	Cry10A/2	

Percentage Divergence

Figure 5.14 Protein sequence relationships between AM1/2, Cry4A, Cry4B and Cry10A/2.
Constructed using a PAM250 residue weight table and manually edited.

From the sequence and phylogenetic analysis a number of interesting questions are raised as to how AM1 and Cry10A should be classified in the new nomenclature which is based entirely on degrees of sequence similarity between δ-endotoxin genes (Crickmore *et al.*, 1998). As previously discussed, classification of Cry10A is based only on the divergent ORF1 region, which is why it has been renamed Cry10A (the only Cry10 gene so far identified) from CryIVC, moving it away from Cry4A and Cry4B in the classification. Presumably, on the same system AM1 will also be classified as a new gene type. However, this classification does nothing to suggest the obvious relationships between Cry10A and AM1 and indeed between these two proteins and Cry4A and Cry4B that have been discussed above. It would be perhaps more accurate to rename Cry10A and AM1, Cry4C and Cry4D, respectively. This classification would be particularly appropriate if the ORF2 regions of these two proteins are found to play a functional role, as has been hypothesied above.

So far, two theories have been discussed as to the evolution of this group of four proteins. Firstly, that a common progenitor of AM1 and Cry10A arose from a 130-kDa protein by the insertion of a DNA fragment. Subsequent divergence of this disrupted protein sequence then led to the evolution of two independent proteins AM1 and Cry10A. Secondly, that a common 130-kDa progenitor of AM1 and Cry10A evolved to form two divergent proteins both of which were independently disrupted by the insertion of DNA fragments.

There is, however, a third theory that goes some way to explaining the intriguing discrepancy in sequence conservation between the N and C-terminal regions of all four proteins. Even in the small section of protein sequence shown in Figure 5.15 the distinct difference in sequence conservation between the two regions can be clearly seen in the distribution of red boxes. This figure theoretically joins the two open reading frames of

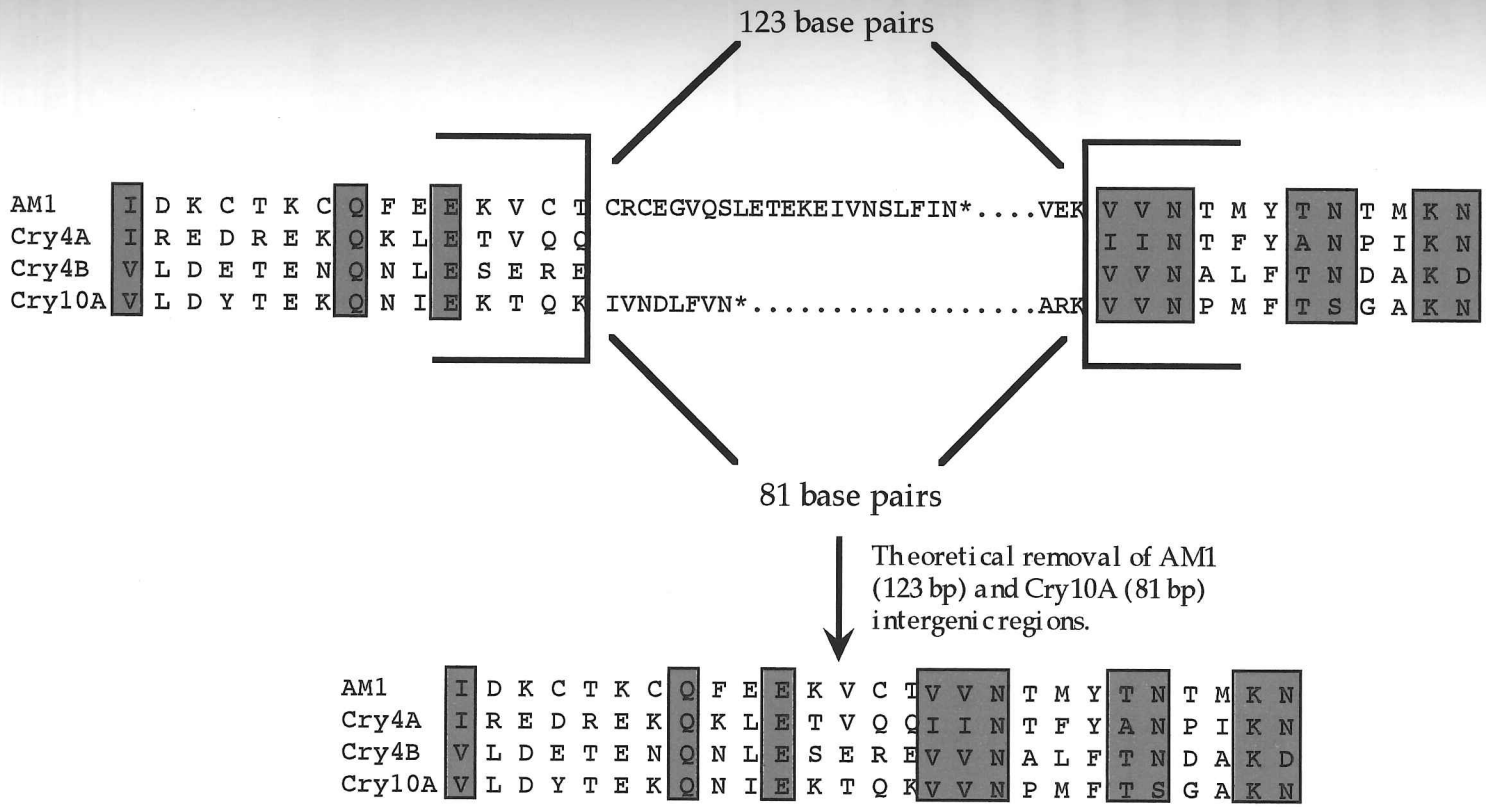


Figure 5.15 Amino acid sequence alignment.
Scheme to show how the removal of the intergenic regions between the two open reading frames of AM1 and Cry10A allows the theoretical construction of complete polypeptide sequences that align with those of Cry4A and Cry4B. Only the relevant sections of sequence are shown. Conserved and semi-conserved residues are boxed in red. Stop codons at the end of AM1 and Cry10A ORF1 are shown as *. The intergenic sequence between the stop codons and the start of the second open reading frame is indicated by dots that do not correspond to the length of the sequence they represent.

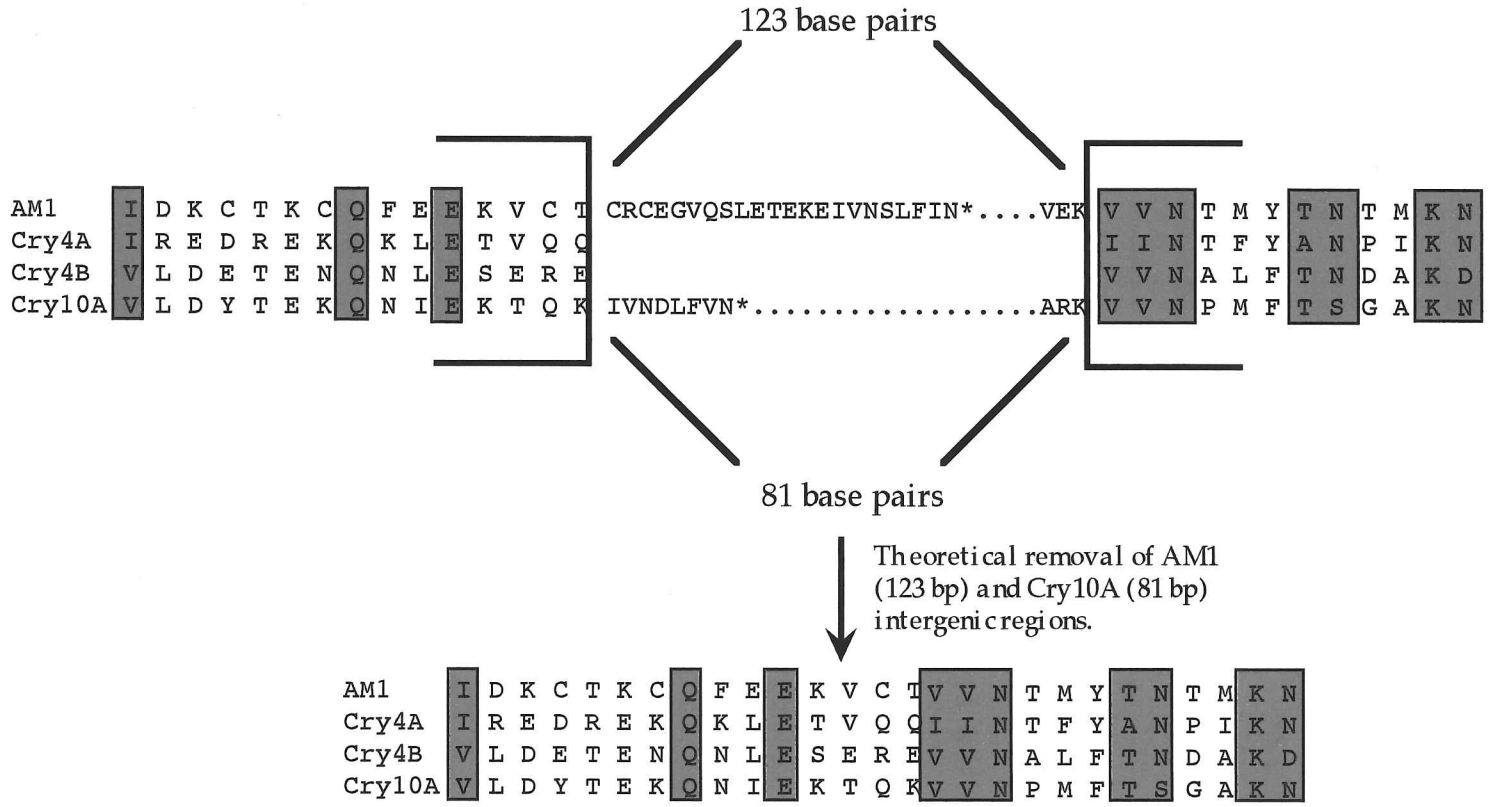


Figure 5.15 Amino acid sequence alignment.
Scheme to show how the removal of the intergenic regions between the two open reading frames of AM1 and Cry10A allows the theoretical construction of complete polypeptide sequences that align with those of Cry4A and Cry4B. Only the relevant sections of sequence are shown. Conserved and semi-conserved residues are boxed in red. Stop codons at the end of AM1 and Cry10A ORF1 are shown as *. The intergenic sequence between the stop codons and the start of the second open reading frame is indicated by dots that do not correspond to the length of the sequence they represent.

Percentage Similarity

	AM1/2	Cry4A	Cry4B	Cry11A	Cry1Aa	Cry1Ba	Cry1Ca	Cry2Aa	Cry3A	
AM1/2		40.3	41.7	12.0	25.3	24.1	25.4	14.2	25.5	AM1/2
Cry4A	53.4		52.5	10.7	24.7	24.9	24.2	11.2	20.0	Cry4A
Cry4B	52.3	40.3		11.0	24.0	27.0	26.9	11.5	17.4	Cry4B
Cry11A	85.9	86.3	85.3		12.6	11.5	13.1	15.2	12.1	Cry11A
Cry1Aa	69.3	65.8	66.8	82.8		55.6	68.4	15.0	28.0	Cry1Aa
Cry1Ba	69.1	66.8	67.2	81.7	40.2		56.6	15.8	32.3	Cry1Ba
Cry1Ca	68.7	66.2	67.3	83.7	28.2	39.0		13.6	27.5	Cry1Cc
Cry2Aa	83.0	85.3	83.0	74.3	80.7	80.1	82.9		14.1	Cry2Aa
Cry3A	68.0	72.6	71.6	83.5	65.4	62.1	63.1	79.1		Cry3A
	AM1/2	Cry4A	Cry4B	Cry11A	Cry1Aa	Cry1Ba	Cry1Ca	Cry2Aa	Cry3A	

Percentage Diversity

Figure 5.16a Protein sequence relationships between AM1/2 and representatives from different classes of δ-endotoxin
Constructed using a PAM250 residue weight table.

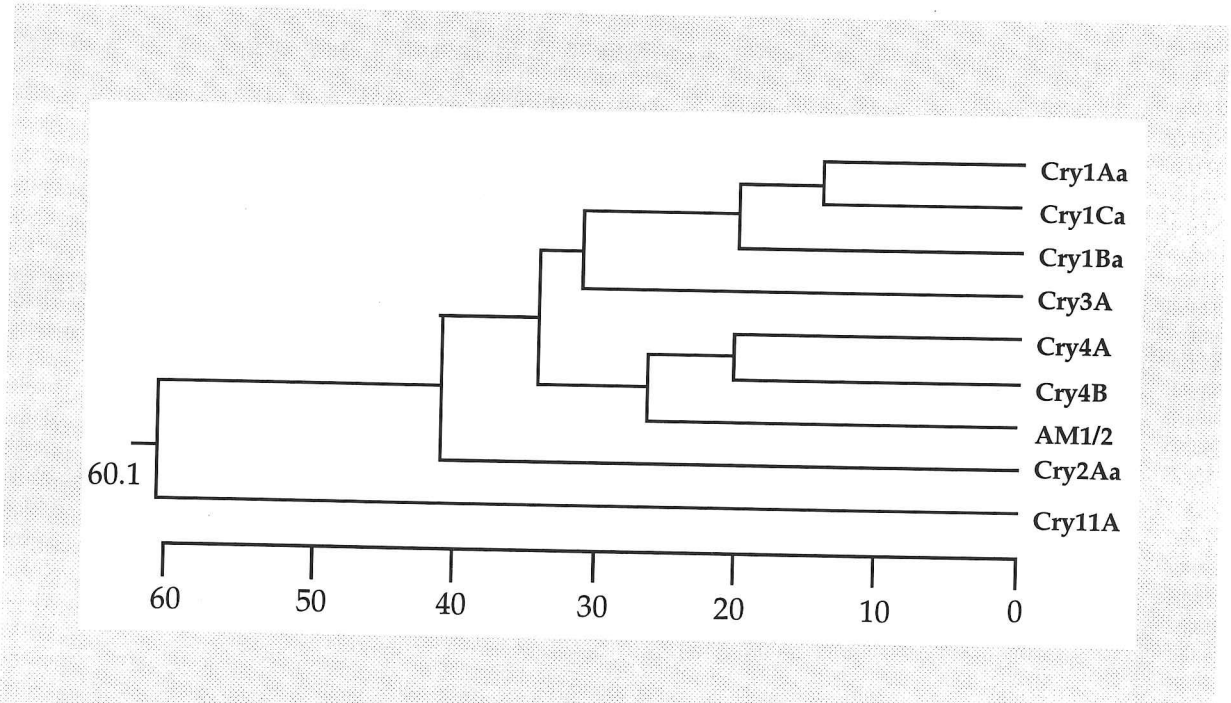


Figure 5.16b Phylogenetic tree based on the sequence relationships of the δ-endotoxins described in Figure 5.16a.
The length of each pair of branches represents the distance between sequence pairs. The scale beneath the tree measures the distance between sequences; units indicate the number of substitition events. The new nomenclature is given in brackets.

AM1 and Cry10A so that they mimic Cry4A and 4B. However, it would almost seem more appropriate to split Cry4A and Cry4B to mimic the two-gene arrangements of AM1 and Cry10A. Based on the level of sequence conservation it is not hard to see a distinct dividing line between what appear to be two different proteins instead of two different regions of the same protein.

This observation mirrors that of a much larger phylogenetic study that has recently been conducted into the whole Cry protein family (Bravo, 1997). In this study, the phylogenetic relationships of the three δ-endotoxin domains were analysed rather than N and C-terminal regions, as has been carried out here, but similar observations were made. Bravo reported (as has been demonstrated above) that the results of phylogenetic analysis of the whole Cry protein sequences do not reflect the complex evolutionary relationships found in the analysis of the independent functional domains. One of the conclusions from this investigation was that the δ-endotoxin domains I, II and III (see Introduction) have evolved independently. The results of phylogenetic analysis of domain I sequences suggest that this domain has a common origin for the whole protein family, while there are three independent origins of domain II and two for domain III. If this theory is extended to AM1 and Cry10A then instead of considering these proteins as disrupted and somehow damaged, these "two-gene" arrangements could be considered as the evolutionary progenitors of "full length" proteins. AM1 and AM2 would therefore represent evolutionary "relics" of the original gene arrangment that became one continuous open reading frame when the intervening region of DNA was lost.

In order to determine to what extent the C-terminal region of AM1/2 corresponded to the domain III discussed in the Bravo study, the amino acid sequence of AM1/2 was aligned with that of the two δ-endotoxin genes for which the crystal structure has already been determined, Cry1Aa and Cry3A (Figure 5.17). Interestingly, the start of AM2 aligned not to the start, but to the end of domain III. As the crystal structure has been determined for the activated toxin, the end of domain III marks the end of the protease-resistant core. It therefore appears that AM1 represents a pre-activated δ-endotoxin protein much like Cry3A. The implications for this are explored in the Discussion to this Chapter.

5.5 Sequencing Strategy for Construct pAM4.4

To determine if this fragment harboured a crystal toxin gene, two subclones of pAM4.4 were constructed. *Pst*I digestion of pAM4.4 results in the production of two restriction fragments, the smaller of which (2.6 kb) hybridises the DIG-labelled 17A.1 probe. Only 1.8 kb of the larger 5 kb fragment consists of 17A plasmid DNA, the remaining 3.2 kb

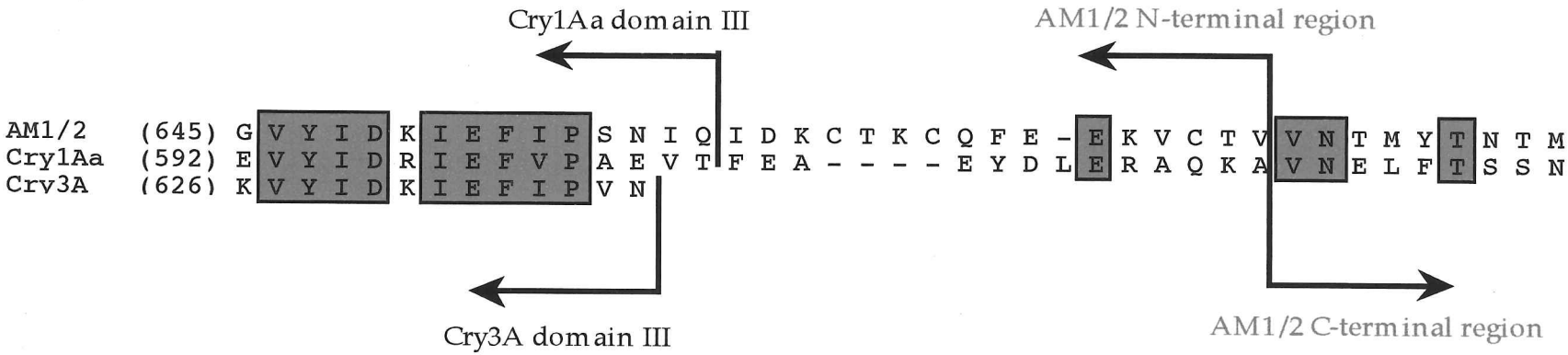


Figure 5.17 Clustal alignment of part of AM1/2 with the corresponding regions of Cry1Aa and Cry3A.
The ends of the recognised domain III regions are indicated for Cry1Aa and Cry3A, as are the N and C-terminal regions of AM1/2. Conserved or semi-conserved sequence is boxed in red. The numbers in brackets indicate the position of the first amino acid in each full length sequence.

being pGEM-3Zf(+) vector. The 5 kb and 2.6 kb *Pst*I fragments were gel-purified as described in Methods. Similarly, pGEM-3Zf(+) was cut with *Pst*I and dephosphorylated. The 5 kb fragment was religated to itself and the 2.6 kb fragment was cloned into the *Pst*I site of pGEM-3Zf(+) and transformed into *E. coli* JM109 by electroporation. Resulting transformants were screened by plasmid DNA minipreparations and restriction digestion. A number of successful subclones were identified.

One subclone representing the 5 kb plasmid (pAMscP1) and one representing the 5.8 kb recombinant (pAMscP2) were selected. Sequence obtained from the ends of each cloned fragment using the T7 and Sp6 primers was obtained by automated sequencing. The DNA and protein sequence was subjected to a BLAST search as previously described, but no similarity to any Bt crystal toxin gene was identified. Less than 1.4 kb of the 2.6 kb hybridising fragment remained unsequenced; this was insufficient DNA to encode a protein of around 70-kDa. However, the remaining DNA was sequenced by "gene walking" as previously described. Analysis of the sequence data obtained did not reveal any similarities to previously identified Bt δ-endotoxins, nor was any region identified with significant similarity to the oligonucleotide probe 17A.1.

5.6 Discussion

The cloning of AM1 represents the second identification of a natural C-terminal truncate of a Cry4-type protein, the first being the 65-kDa Cry10A from Bt ssp. *israelensis* (Bti) (Thorne *et al.*, 1986). The missing C-terminal region of both proteins appears to be encoded as a complete Orf immediately downstream of the first protein gene.

Irrespective of these similarities, AM1 and Cry10A appear to be no more closely related to each other than they are to the full-length Cry4A and Cry4B, indicating that all four proteins diverged from a common ancestor at a similar point in evolutionary history. Cry4A has been found to be associated with transposable elements (Delécluse *et al.*, 1989) and it has been reported that a member of the IS*240* family of Insertion Sequences is present in evey diptericidal strain (Delécluse *et al.*, 1997). The presence of these transposable elements may therefore provide a mechanism for replication of a δ-endotoxin gene, copies of which will then be subject to different substitutional events, giving rise to a family of different but related proteins. As the majority of δ-endotoxin genes have also been found to be plasmid borne this also provides a means of transferring δ-endotoxin gene copies to other Bt strains. The level of sequence divergence between the four proteins studied here indicates either that AM1 should be considered as a novel toxin gene, or that Cry10A has been wrongly classified in the new nomenclature and Cry10A and AM1 should be renamed Cry4C and Cry4D, respectively.

The breaks between the two Orfs of AM1 and Cry10A occur at almost exactly the same place, raising the question as to whether these "two-gene" arrangements result from the disruption of a Cry4-type 130-kDa protein gene, or are in fact the evolutionary progenitor of the "full-length" proteins. This latter arrangement would support Bravo's recent phylogenetic study (Bravo, 1997) where it was observed that the different functional domains of a Bt toxin had evolved independently.

To "reconstruct" 130-kDa proteins theoretically from the two open reading frames of AM1 and Cry10A it was necessary to remove a small region (123 bp for AM1, 83 bp for Cry10A) of intergenic DNA. This process indicates that an insertion (or deletion) of a DNA fragment was responsible for the observed difference in the gene arrangements of the 65 and 130-kDa protein types, not a series of point mutations.

Alignments also revealed marked differences in the level of sequence conservation between the N and C-terminal regions. The N-terminal regions of all four proteins showed similarities of between about 25-35% and appeared to have originated from a common ancestor at a very similar point in evolutionary history. The level of sequence conservation in the C-terminal regions was much more marked, lending further support to the theory that the N and C-terminal regions had evolved from different origins. The sequence similarity between Cry4A and Cry4B C-terminal regions was far higher (>90%) than that shown between either of these proteins and AM1 or Cry10A (55-65%) or between AM1 and Cry10A themselves suggesting that this region may be functionally conserved. The lower level of sequence conservation in the dissociated C-terminal regions of AM1 and Cry10A further supports this theory.

However, it is also hypothesised in this Chapter that the dissociated C-terminal regions of AM1 and Cry10A may serve some functional purpose. Otherwise it is unclear why they should have retained their integrity as open reading frames and show such a high degree of sequence conservation. ORF 2 of both of these proteins seems to correspond to the region that is removed during protoxin activation. This protein has been proposed to be important in stabilising and packaging the polypeptide chains into the crystal (Choma *et al.*, 1991). Cry10A ORF2 has been shown to be expressed and is present in the native Bti inclusion (Purcell, 1997) perhaps indicating that it plays some stabilising role. It will be interesting to see if AM2 is similarly expressed.

Whether the two-gene arrangements of AM1 and Cry10A are the progenitor or the progeny of the 130-kDa proteins is not clear. However, it seems intriguing that in the case of Bti at least, copies of both arrangements have been retained. Thorne *et al.* (1986) proposed that the existence of related but different toxin genes in a single isolate afford the cell the luxury of a more rapid mutation frequency, while not risking the elimination of all toxin activity in one step. The result could be a potential broadening of the spectrum of insecticidal activity for a given strain. It is also probable that the maintenance of several

different variants of the same gene postpone the development of insect resistance (Van Rie *et al.*, 1990b).

In this Chapter, AM1 has been likened to the "protease-resistant core" of a longer δ-endotoxin, much like Cry3A. Thorne *et al.* (1986) made similar comparisons between the 130-kDa protein of Bt ssp. *kurstaki* and Cry10A of Bti and noted that the ORF1 component of Cry10A resembled the region within the Bt ssp. *kurstaki* gene that coded for a protease-resistant core with elevated insecticidal specific activity. It was hypothesised that the synthesis of a shortened toxin by Bti might eliminate the need for proteolytic activation in the mosquito larval gut. This might in turn extend the spectrum of activity of this strain to more species of mosquitoes. Purcell (1997) has since found Cry10A to be non-toxic to *Aedes aegypti* larvae, and that it was necessary to add a C-terminal region from Cry4B before any activity was shown against this insect. However, Cry10A was only tested against *Aedes aegypti* larvae, a species highly susceptible to Cry4A and Cry4B whereas the Thorne *et al.* (1986) hypothesis predicts that Cry10A might show activity against insect species not affected by Cry4A and 4B. Bti is also toxic to blackfly larvae, so it is possible that Cry10A may play a role in this or other diptericidal activities.

Extending this hypothesis to the newly identified AM1, it is possible that this essentially pre-activated δ-endotoxin is more active against susceptible dipteran larvae (*Dacus oleae*) than a full length toxin requiring proteolytic activation. However, the C-terminal protein region may still be required to assist packaging into the inclusion and to provide a stable, protease-resistant crystal. This could explain the survival of *am2* as a complete open reading frame and its sequence conservation to other C-terminal regions.

Therefore, the existence of a Cry4-type protein as either a two-gene or single-gene arrangement may be dependent on the gut conditions (more specifically the protease conditions) of the suseptible insect species. In some species, e.g. *Aedes aegypti*, a high level of specific and non-specific protease acitivity may mean that activiation of the protoxin is very efficient and that degradation of the δ-endotoxin is rapid. The added stability of an attached C-terminal region may therefore be essential. In other species, e.g. *Dacus oleae*, δ-endotoxin proteolysis may proceed at a reduced rate. The existence of a pre-activated toxin may therefore provide an adaptational advantage to the Bt species in environments where degradation of the toxin is not a limiting factor.

This hypothesis is in effect dependent upon an equilibrium between (i) the advantage of possessing a pre-activated "ready-to-go" toxin and (ii) the disadvantage of the toxin being degraded before it can exert an effect. This balance is affected by the gut conditions of the susceptible insect. A study into how gut conditions such as protease content and pH effect the relative susceptibility of Lepidopteran species to certain δ-endotoxins is currently being carried out in this laboratory (Kaile-Gilliland *et al.*, 1997; Chambers *et al.*, 1997) and may eventually provide an insight into how these conditions might also affect diptericidal δ-

endotoxins. As other diptericidal Bt strains are investigated and further Cry4 and Cry10-type proteins are characterised it may be possible to build a clearer picture of the evolutionary origins of these genes.

In conclusion, the four proteins, Cry4A, Cry4B and Cry10A from Bt ssp. *israelensis* and the newly identified AM1 from Bt ssp. *fukuokaensis* 17A have considerable primary sequence similarity.

The smaller AM1 and Cry10A may represent pre-activated variants of the longer proteins. The remaining C-terminal regions of these proteins are encoded as complete open reading frames, the gene products of which may be present in the native inclusion to stabilise the crystal and help prevent proteolysis. The co-existence of two-gene and single-gene arrangements of this same δ-endotoxin family may be as a result of adaptations to different gut conditions in suceptible insects.

130-kDa proteins were theoretically "reconstructed" from the two Orfs of AM1 and Cry10A by the removal of a small intergenic region of DNA, leading to the conclusion that an insertion (or deletion) of a DNA fragment was responsible for the different gene arrangements.

The question remains as to whether the single-gene arrangement is the evolutionary progenitor of the two-gene arrangement or *vice versa*. A study carried out by Bravo (1997) concluded that the different δ-endotoxin domains evolved independently which would support the latter theory.

It is possible that Cry10A is an evolutionary "relic", surplus to requirements in a strain with highly adapted and active 130-kDa proteins. However, this would not appear to be the case for AM1 as no 130-kDa proteins were identified in the polypeptide profile. It therefore remains to be determined, how much of a role this protein plays in the toxicity of 17A towards the Olive fruit fly. It would also be interesting to discover if *am2* is expressed, as this would lend weight to one or possibly two theories: firstly, that the C-terminal region of the 130-kDa Cry10A type proteins originated from an independently expressed and functional protein moiety, and secondly, that in the shorter proteins the second open reading frames have been functionally conserved.

To address these issues the following Chapter investigates expression of *am1* and *am2*.

Six

Chapter 6

Expression of AM1 in Bacillus thuringiensis

6.1 Introduction

Over 140 δ-endotoxin genes have been cloned from different Bt subspecies and many of their gene products obtained by expression in *E. coli*, *B. subtilis* (Whiteley & Schnepf, 1986) or an acrystaliferous strain of Bt (Crickmore & Ellar, 1992). In Bt, cloned genes as for their native counterparts, are generally only transcribed during sporulation (Brown & Whiteley, 1988, 1990). The high level of crystal protein expression is believed to be due to the stability of the mRNA produced during sporulation (Petit-Galtron & Rapoport, 1976) and the presence of strong Bt promoter sequences. In some cases, these promoters are used sequentially during sporulation allowing transcription over a prolonged period (Brown & Whiteley, 1988, 1990). Expression of δ-endotoxins in *E.coli* has usually involved constructs containing inducible promoters such as *lacZ* (Angsuthanasombat *et al.*, 1987) or *tac* (Smith, 1992) to yield high levels of δ-endotoxin which accumulates as inclusion bodies.

This Chapter describes the expression of the AM1 gene previously cloned from Bt ssp. *fukuokaensis* 17A in IPS78/11. This strain is a crystal minus strain of Bt ssp. *israelensis* containing only two native plasmids and is commonly used in this laboratory for expression purposes. Bt rather than *E. coli* was selected as the recipient as this environment would more closely resemble that of *fukuokaensis*. Using a Bt strain should also ovecome the potential problem of promoters and it was therefore hoped that, using this system, expression could be studied from the genes own promoter.

Before AM1 could be expressed in Bt it was necessary to subclone the gene into a shuttle vector containing a Bt origin of replication as well as one for *E coli*, thereby enabling replication of the foreign DNA in both hosts. Attempts were also made to 'repair' the lesion between *am1* and *am2* by overlap extension PCR, in order to create a novel 130-kDa crystal protein.

6.2 Expression of *am1* in Bt ssp. *israelensis* IPS78/11

The Bt shuttle vector pSVP27A is the vector of choice for expression of Bt genes in IPS78/11 (Crickmore *et al.*, 1990; Crickmore & Ellar, 1992). pSVP27A is a derivative of the shuttle vector pSV2 which itself is a chimera of pBR322, pUC18 and pC194 containing a

700 bp *Eco*RV-*Hae*III fragment from the upstream region of the *cyt1A* gene including the highly efficient *cyt1A* promoter. The presence of this region provides the option of expressing the cloned gene either under its own, or the *cyt1A* promoter depending upon the orientation of the cloned fragment.

Initially, a subclone of the first available construct, pAM15, was used from which to express *am1*. A 4 kb *Pst*I fragment containing *am1* and its upstream promoter region was selected, but lack of suitable restriction sites in the pSVP27A polylinker meant that an intermediate stage in pGEM-3Zf(+) had to be utilised. The 4 kb *Pst*I fragment from pAM15 was purified from a 1x TAE 0.8% agarose gel by GeneClean II (see section 2.2.5) and cloned into *Pst*I cut, dephosphorylated pGEM-3Zf(+) purified in the same way as the insert DNA. A 3:1 ratio of insert to vector DNA (as determined from an ethidium bromide stained gel) was ligated in an overnight reaction at 15°C as described in section 2.2.8. 1 µl of the ligation mixture was transformed into *E. coli* JM109 and transformants selected on LB plates containing 100 µg/ml ampicillin. The plasmid content of the colonies was analysed by small scale DNA preparation, restriction endonuclease digestion and agarose gel electrophoresis. Nine positive clones were obtained out of ten colonies analysed. One of which was found by restriction analysis to have *am1* orientated in the anticlockwise direction (pGEM*Pst*I-).

This orientation was selected, as it placed the *Sph*I sites from the old and new vector molecules at opposite ends of the cloned fragment. This enabled the subsequent removal of a 4 kb *Sph*I fragment containing all of the insert DNA from pGEM*Pst*I-. The *Sph*I fragment was gel purified then cloned into the *Sph*I site of the 5.8 kb shuttle vector pSVP27A in a procedure similar to that detailed above. Of the ten transformant colonies analysed one positive clone (pSVP27AM1-) was identified. The presence of the *am1* gene in this clone was confirmed by hybridisation using the DIG-labelled 17A.1 probe (Sections 2.2.10-2.2.13). This clone was found, by restriction digest analysis to have *am1* orientated in the anticlockwise direction. This cloning strategy is outlined in Figure 6.1. As can be seen from Figure 6.1, this orientation of the insert does not place *am1* under the influence of the *cytA* promoter. This was desirable so that expression of *am1* could be investigated from its own promoter.

pSVP27AM1- DNA was used to transform IPS78/II (Section 2.2.16b) and transformants selected on LB agar containing 10 µg/ml chloramphenicol. The presence of the correct plasmid in these clones was verified by restriction analysis and hybridisation using the DIG-labelled 17A.1 probe. Expression of the *am1* gene on the pSVP27AM1- plasmid in IPS78/11 was induced by streaking plasmid containing colonies onto a medium designed to induce rapid sporulation (CCY agar containing 10 µg/ml chloramphenicol).

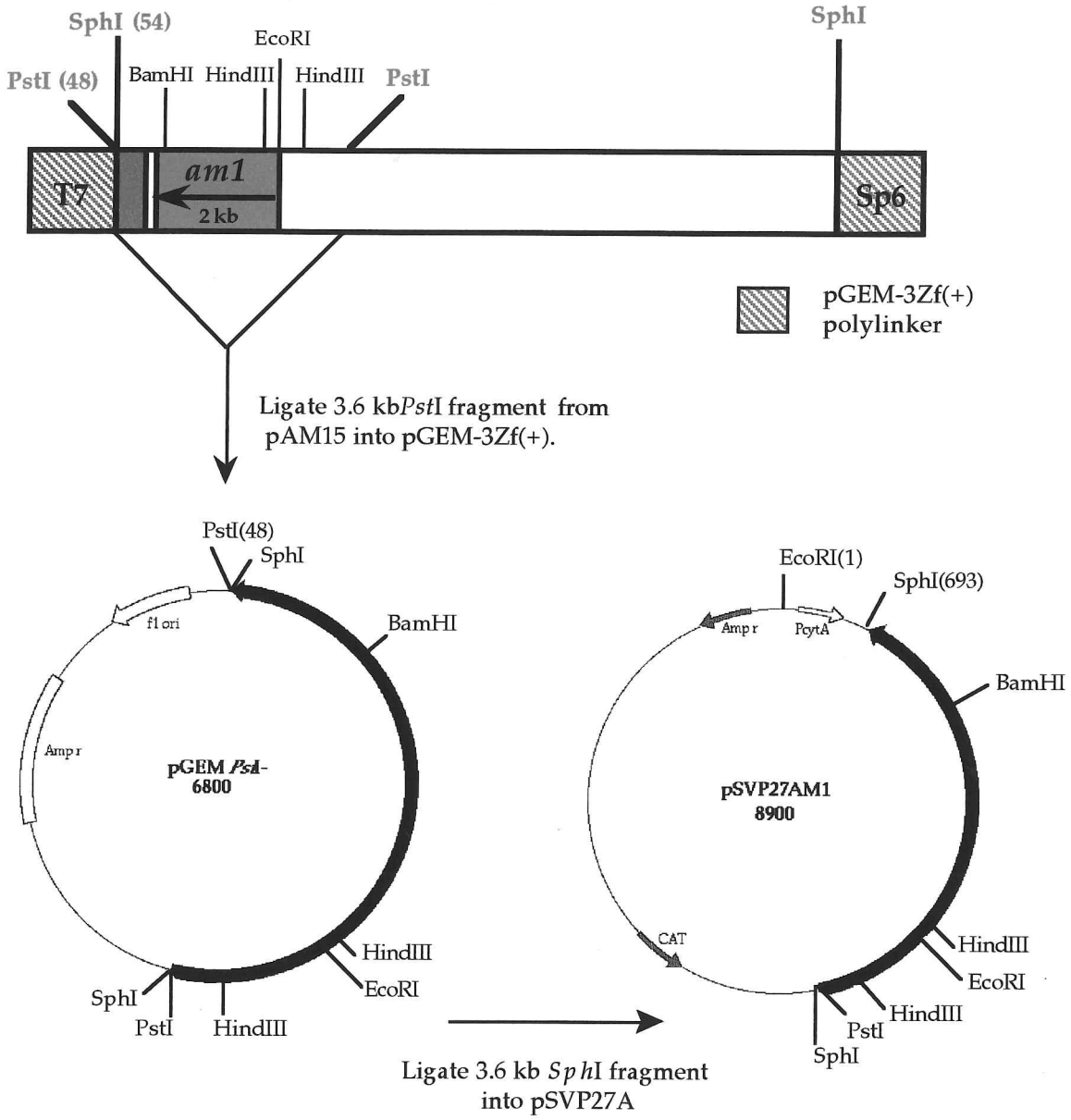


Figure 6.1 Subcloning of a 3.6 kb *Sph*I fragment containing *am1* and the incomplete open reading frame *am2* into the Bt shuttle vector pSVP27A.
The construct pAM15 is shown as a bar map. The positions of the T7 and Sp6 priming sites are indicated. The position of *am1* is indicated by a red box and that of the incomplete open reading frame *am2* by a purple box.
Constructs pGEM*Pst*I- and pSVP27AM1 are represented by vector circle maps, but are not to scale.
The thick black arrow indicates the direction of *am1* transcription in both cases.

After approximately 24-36 hours the cells were observed (microscopically) to have sporulated and the presence of crystals was investigated. Very small, phase dark, spherical inclusion bodies were observed in some, but not all cells. It was believed that these could contain AM1 crystal protein and it was decided to analyse these inclusions further.

The recombinant IPS78/11 was grown in 1 litre of CCY liquid medium in the presence of 10 µg/ml chloramphenicol at 30°C. Complete lysis took place after 36 hours. The inclusions were separated by a sucrose density gradient (Section 2.3.1) and a very thin grey band was observed between the interface of the 1.97 and 2.10 M sucroses. The material from this band was removed and washed several times with ice-cold water. Observation of the preparation by phase-contrast microscopy revealed the presence of a low concentration of the very small, phase dark, spherical inclusion bodies that had initially been observed.

After purification, the inclusions were solubilised in gel sample buffer containing SDS and subjected to SDS-PAGE analysis (Section 2.3.4). It was found that 1/5th of the total preparation (from 1 L of culture) had to be concentrated and run on the gel before any band could be seen. The profile of this band was compared to that of whole crystals prepared from the parent Bt ssp. *fukuokaensis* 17A strain, harbouring the *am1* gene (Figure 6.2). One faint band was observed, which at about 37-kDa did not correspond to any of the 17A proteins. The remaining portion of preparation was concentrated in an attempt to obtain enough material for N-terminal sequence analysis (Section 2.3.8) but this did not prove to be possible. The identity of this 37-kDa protein has not therefore been determined.

The possibility that the recombinant strain had lost the plasmid encoding *am1* was ruled out as restriction digest analysis of a small scale DNA preparation taken from the culture confirmed the presence of pSVP27AM1-. The presence of chloramphenicol in the growth media should also ensure plasmid maintenance. It was therefore concluded that AM1 is not expressed from this construct.

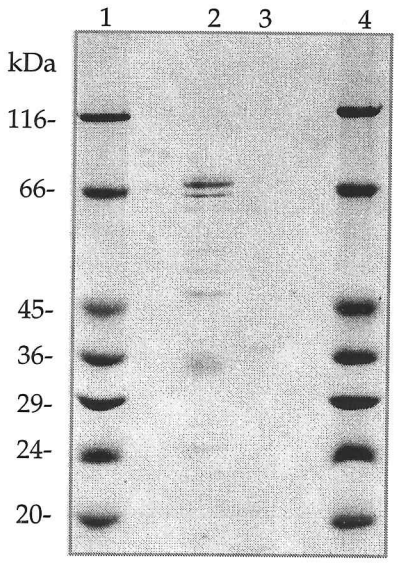


Figure 6.2 Coomassie blue stained 13% SDS-PAGE analysis of sucrose gradient purified inclusions from Bt ssp. *fukuokaensis* 17A (lane 2) and Bt IPS78/11 harbouring PSVP27AM1- (lane 3). Molecular weight markers are shown in lanes 1 and 4.

This lack of expression is perhaps not surprising given that *am1* appears to lack a functional C-terminal region. As discussed in Chapter 5, in the full length Cry4A and 4B proteins this C-terminal region is believed to act as a form of internal chaperone, either by directing toxin folding to a stable conformation or by protecting the unfolded toxin from proteolytic degradation.

The best studied example of a molecular chaperone is the effect of a 20-kDa protein from Bti on the production of Cyt1A protein. Cyt1A is produced at very low levels in Bt (Crickmore, personal communication) and *E. coli* (McLean & Whiteley, 1987; Visick & Whiteley, 1991; Adams *et al.*, 1989) unless this 20-kDa protein (encoded by *p20*) is present. *p20* is situated 4 kb upstream of the *cyt1A* gene (see Figure 6.3). When these two Orfs were cloned onto the same fragment, the introduction of nonsense codons into *p20* abolished the enhancing effect confirming its function (Adams *et al.*, 1989). Gene specific mRNA levels were measured and found to be the same both in the presence and absence of the 20-kDa protein indicating that the effect of the latter is post transcriptional (McLean & Whiteley, 1987; Adams *et al.*, 1989). Further work suggested that the rate of translation was also unaffected (Visick & Whiteley, 1991) indicating that the 20-kDa protein must act on the Cyt1A protein, possibly by having a stabilising effect rather than by regulating gene expression parsec.

The 20-kDa protein was then found to be dispensable for high level Cyt1A production in mutants with reduced proteolytic ability (Visick & Whiteley, 1991). It was proposed that since any full length, fully folded Cyt1A that is synthesised is relatively stable, the target for proteolytic attack might be the unfolded or nascent peptide. Cyt1A and the 20-kDa protein were co-immunoprecipitated from *E. coli* extracts by antibodies against either peptide (Visick & Whiteley, 1991), which may be indicative of a protein-protein interaction between the two. This hypothesis was further supported by the finding that co-immunoprecipitation occurred only when both proteins were made in the same cell and not when extracts were mixed post synthetically. This suggests that the 20-kDa protein is only able to bind Cyt1A during Cyt1A synthesis and it is therefore thought that the 20-kDa protein protects the unfolded Cyt1A peptide from proteolytic attack and perhaps also directing its folding.

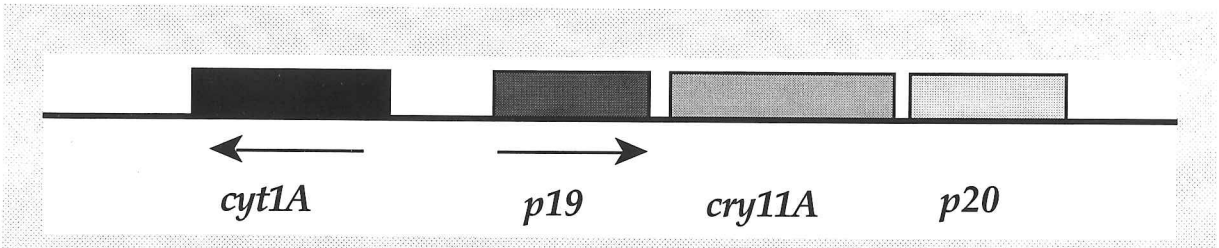


Figure 6.3 Schematic representation of the genes present in the *cry11A* operon from Bt ssp. *israelensis*.
Arrows indicate the direction of transcription as determined by Dervyn *et al.*, 1995.

Expression of Cry11A has also been found to be significantly enhanced in *E.coli* by the presence of the 20-kDa protein (Visick & Whiteley, 1991). Cry11A like Cry10A has a significantly reduced C-terminal region, it is therefore likely that both polypeptides will require some form of chaperone for efficient expression.

Even though a chaperone may necessary for efficient crystallisation of AM1 is was thought possible that AM1 protein may be present in a soluble non-crystalline form within the cell. To investigate this possibility recombinant IPS78/11 culture was treated with 15% tri-chloroacetic acid (TCA) which precipitates all soluble proteins. Samples of culture were taken both before and after cell lysis as proteins are more likely to be proteolysed after cell lysis. 1.5 ml of whole, pellet and supernatant fractions of each sample were lysed by sonication, TCA precipitated and resuspended in 20 µl (Section 2.3.9-2.3.10). The full 20 µl was then solubilised in gel sample buffer and analysed by SDS-PAGE alongside crystal protein from 17A and an IPS78/11 negative control (Figure 6.4) but no protein corresponding to expression from *am1* was observed.

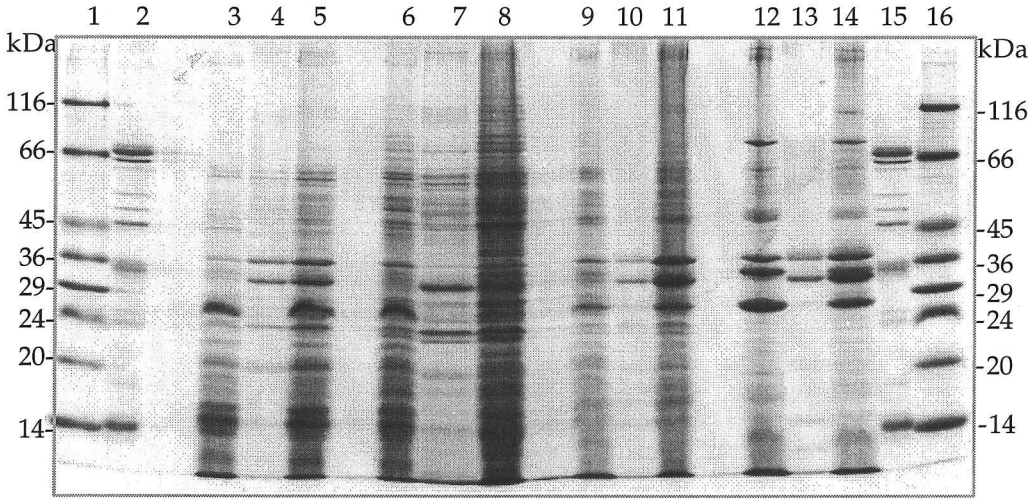


Figure 6.4 Coomassie blue stained 14% SDS-PAGE gel of TCA precipitates of IPS78/11 (pSVP27AM1-) from culture taken at the following times after inoculation:-

Lane 1	Molecular weight markers
Lane 2	10 µg of 17A crystals
Lanes 3,4 &5	14 hour culture, whole, supernatant and pellet fractions
Lanes 6,7&8	18 hour culture, whole, supernatant and pellet fractions
Lanes 9,10&11	24 hour culture, whole, supernatant and pellet fractions
Lanes 12,13&14	24 hour culture of IPS78/11, whole, supernatant and pellet fractions
Lane 15	10 µg of 17A crystals
Lane 16	Molecular weight markers

The observation that the gene product of the *cry10A* ORF2 was present in Bti inclusions (Purcell, 1997) raised the possibility that *am2* might be likewise expressed. It was therefore hypothesised that an independently expressed C-terminal region may act as a molecular chaperone in much the same way as the attached C-terminal regions of Cry4A and 4B and enhance expression of AM1. In order to determine if AM2 could be expressed work was started to construct a second clone that would harbour DNA for the complete open reading frames of *am1* and *am2*. The construction of this clone is detailed in Chapter 5. Meanwhile, it was decided to investigate expression of *am1* under the influence of the *cyt1A* promoter present in the shuttle vector pSVP27A.

In order to study *am1* expression from the *cyt1A* promoter the orientation of the cloned insert in pSVP27AM1- had to be reversed. pSVP27AM1- DNA was digested with *Sph*I and the two resulting fragments separated by gel electrophoresis, recovered by Geneclean II, then religated according to the conditions detailed in Section 2.2.8. 1 µl of the ligation mix was transformed into electrocompetent JM109 (Section 2.2.16a) and successful clones identified by restriction analysis of plasmid DNA from transformant colonies. Of eight colonies analysed, one harboured a clone with *am1* in the clockwise orientation (pSVP27AM1+). A restriction map of pSVP27AM1+ is shown in Figure 6.5.

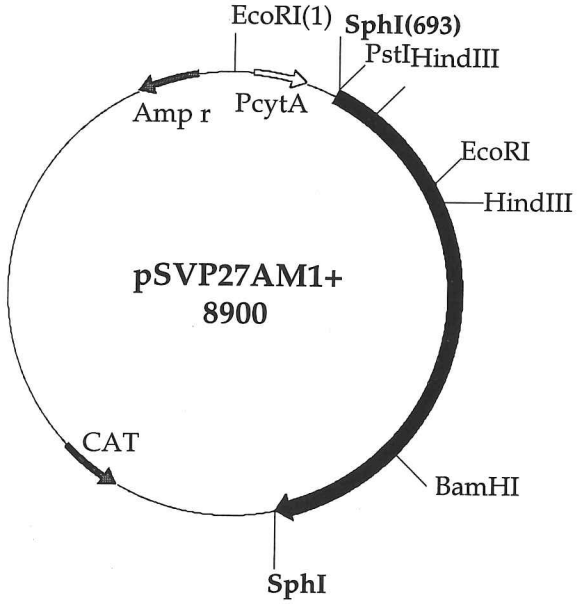


Figure 6.5 Restriction map of plasmid construct pSVP27AM1+.
The 3.6 kb fragment (black box) is cloned into the *Sph*I site of the Bt shuttle vector pSVP27A. The ampicillin and chloramphenicol resistance genes are represented as Amp r and CAT respectively. The thick black arrow indicates the direction of *am1* transcription. The plasmid pSVP27A is a derivative of the shuttle vector pSV2 which is itself a chimera of pBR322, pUC18 and pC194. It contains a 700 bp *Eco*RV-*Hae*III fragments from the upstream region of the *cyt1A* gene including the *cyt1A* promoter (PcytA) (Crickmore *et al.*, 1990; Crickmore & Ellar, 1992).

pSVP27AM1+ DNA was used to transform IPS78/11 (Section 2.2.16b). and transformants selected on LB agar containing 10 µg/ml chloramphenicol. The presence of the correct plasmid in these clones was verified by restriction analysis and hybridisation using the DIG-labelled 17A.1 probe. Expression of *am1* was induced by streaking plasmid containing colonies of IPS78/11 onto CCY agar containing 10 µg/ml chloramphenicol. After approximately 24-36 hours the sporulating cells were observed (microscopically) to contain dark, irregular inclusion bodies of much greater size and quantity than those observed from IPS78/11(pSVP27AM1-).

IPS78/11(pSVP27AM1+) was grown in CCY liquid medium in the presence of 10 µg/ml chloramphenicol at 30°C. Cell lysis was complete after about 36 hours. The inclusions were separated from the spores and cell debris in a discontinuous sucrose gradient (1.97 M, 2.10 M, 2.30 M) as described in Section 2.3.1. Two bands were observed one at the 1.97-2.1 M interface (top fraction) and one at the 2.1-2.3 M interface (bottom fraction). Both bands and the pellet material were removed separately and washed with ice-cold water. The purity of the three preparations was assessed by viewing the sample by phase-contrast microscopy. The top fraction was found to be composed of about 95% crystals, 5% spores and the bottom fraction about 90% crystals, 10% spores. The pellet was composed mainly of spores. Crystal protein recovery was determined by the Lowry method (Section 2.3.3). Both fractions were found to contain approximately 1 g of protein per litre of CCY culture.

After purification, the inclusions were solubilised in gel sample buffer containing SDS and subjected to SDS-PAGE analysis. The profile of the sucrose density purified crystals was again compared to that of whole 17A crystals (Figure 6.7). Two major bands were noted of approximately 70 and 25-kDa.

The top band seems to correspond to the 70-kDa band of 17A and to not the 65-kDa band. This discrepancy in size (and the discrepancy in N-terminal sequence as determined from the DNA sequence in Chapter 5) may well indicate that the oligonucleotide probe designed against the 65-kDa protein of 17A has in fact hybridised to the gene for a larger protein with a similar N-terminus. The presence of such a gene was indicated in the mixed sequence results obtained from N-terminal sequencing of the 17A 70-kDa band (see Chapter 3) and also by the presence of more than one hybridising band in each native plasmid restriction digest (see Figure 4.5).

The presence of the lower 25-kDa band was also intriguing. There is just enough *am2* DNA present in construct pSVP27AM1+ to encode a protein of approximately 28-kDa. It was therefore conceivable that this highly truncated open reading frame had been expressed and further truncated to form a protein of 24-kDa. It was of course also possible that this band was a degradation product of the 70-kDa protein.

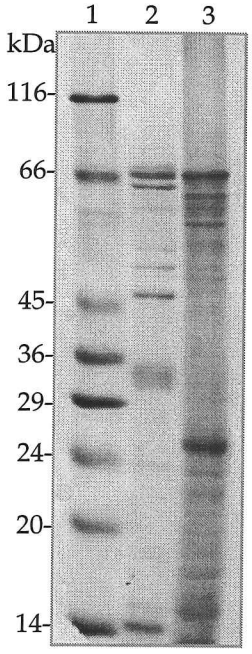


Figure 6.6 Coomassie blue stained 13% SDS-PAGE analysis of sucrose gradient purified inclusions from Bt IPS78/11 harbouring pSVP27AM1+ (lane 2) and Bt ssp. *fukuokaensis* 17A (lane 3). 10 µg of both samples were solubilised for 5 minutes at 100°C in SDS-PAGE sample buffer. Molecular weight markers are shown in lane 1.

In order to investigate the identities of these protein bands they were blotted onto Problot membrane and N-terminally sequenced (Section 2.3.8). The 70 and 24-kDa bands were found to have the following N-terminal sequences.

70-kDa MNSYQNKNEYEILDASRNNS

24-kDa MYTNTMKNKLKIETTDYEID

The 70-kDa protein had an identical N-terminus to that of AM1 deduced from its DNA sequence in Chapter 5. The expression of a 70-kDa protein from a gene encoding one of nearly 80-kDa probably results from endogenous proteolysis of the full-length protein by the producer organism IPS78/11, a phenomenon that also appears to occur in 17A. From the available information it does appear that a protein from the 70-kDa band of 17A has been cloned and not the 65-kDa protein. It cannot yet be ruled out however that there was a single mistake in the original sequence analysis (a K for an N) and that the 80-kDa gene product is processed differently by 17A than by IPS78/11 to generate a 65-kDa end product. It may also be the case that two populations of end products are present in 17A (one at 70-kDa and one at 65-kDa) only the 70-kDa one of which is present in IPS78/11.

The 24-kDa protein had an identical N-terminus to that of AM2 as deduced from its DNA sequence in Chapter 5, confirming that this highly N-truncated protein has been expressed. It is unusual for a protein as incomplete as this to be expressed in such a stable form. The presence of a stable 24-kDa product of this potentially 60-kDa protein may well

indicate that it is unusually resistant to proteolysis providing further evidence of its role as a chaperone for AM1.

6.3 Generation of Clones Suitable for Expressing *am1* and *am2* in Bt ssp. *israelensis* IPS78/11

A large clone, pAM10, was now available containing the whole of the *am2* open reading frame (see Chapter 5). It was intended that a smaller region containing *am1* and *am2* should be subcloned in the shuttle vector pSVP27A in order to express both open reading frames and further investigate the protein products of the two genes. No suitable sites were available, so they were created using PCR as detailed below. In parallel with the construction of this clone, attempts were also made to remove the intergenic segment of DNA between *am1* and *am2* in such a way that one continuous open reading frame would be formed. It was hoped that this would then enable the expression of a new 130-kDa crystal protein that may have novel toxicities.

A similar procedure had been successfully carried out on *cry10A* from Bti (Zhang, personal communication). In this work the truncated *cry10A* gene was 'repaired' by the addition of the *cry4B* C-terminal region. Cry10A inclusions were found to be non-toxic against 3-day old *Aedes aegypti* larvae. However, the hybrid protein comprised of the naturally truncated Cry10A with an additional C-terminal tail from Cry4B was found to be weakly mosquitocidal (Purcell, 1997).

The similarity in the genetic arrangement of the two Orfs of AM1 and Cry10A has already been discussed a length in Chapter 5. The construction of the *cry10A/cry4B* chaemeric gene was facilitated by conveniently placed restriction sites. No such sites existed between *am1* and *am2*. Sites could have been engineered using PCR however, it was decided to try a different technique, that of overlap extension PCR (Steffan *et al.*, 1989). In this technique, complementary oligonucleotide primers and the polymerase chain reaction (PCR) are used to generate two DNA fragments having overlapping ends. These fragments are then combined in a subsequent 'fusion' reaction in which the overlapping ends anneal, allowing the 3' overlap of each strand to serve as a primer for the 3' extension of the complementary strand. The resulting product is amplified further by PCR. The requirement for restriction endonucleases or DNA ligases is therefore eliminated by generating two PCR fragments with overlapping ends.

This method is illustrated in Figure 6.7. In separate PCRs two fragments of the target gene sequence are amplified (*am1* and *am2*). Each reaction uses one flanking primer that hybridises to DNA at one side of the target sequence (primer **a** or **d** in Figure 6.7) and one internal primer that hybridises to one side of the intergenic region to be removed (primer

indicate that it is unusually resistant to proteolysis providing further evidence of its role as a chaperone for AM1.

6.3 Generation of Clones Suitable for Expressing *am1* and *am2* in Bt ssp. *israelensis* IPS78/11

A large clone, pAM10, was now available containing the whole of the *am2* open reading frame (see Chapter 5). It was intended that a smaller region containing *am1* and *am2* should be subcloned in the shuttle vector pSVP27A in order to express both open reading frames and further investigate the protein products of the two genes. No suitable sites were available, so they were created using PCR as detailed below. In parallel with the construction of this clone, attempts were also made to remove the intergenic segment of DNA between *am1* and *am2* in such a way that one continuous open reading frame would be formed. It was hoped that this would then enable the expression of a new 130-kDa crystal protein that may have novel toxicities.

A similar procedure had been successfully carried out on *cry10A* from Bti (Zhang, personal communication). In this work the truncated *cry10A* gene was 'repaired' by the addition of the *cry4B* C-terminal region. Cry10A inclusions were found to be non-toxic against 3-day old *Aedes aegypti* larvae. However, the hybrid protein comprised of the naturally truncated Cry10A with an additional C-terminal tail from Cry4B was found to be weakly mosquitocidal (Purcell, 1997).

The similarity in the genetic arrangement of the two Orfs of AM1 and Cry10A has already been discussed a length in Chapter 5. The construction of the *cry10A/cry4B* chaemeric gene was facilitated by conveniently placed restriction sites. No such sites existed between *am1* and *am2*. Sites could have been engineered using PCR however, it was decided to try a different technique, that of overlap extension PCR (Steffan *et al.*, 1989). In this technique, complementary oligonucleotide primers and the polymerase chain reaction (PCR) are used to generate two DNA fragments having overlapping ends. These fragments are then combined in a subsequent 'fusion' reaction in which the overlapping ends anneal, allowing the 3' overlap of each strand to serve as a primer for the 3' extension of the complementary strand. The resulting product is amplified further by PCR. The requirement for restriction endonucleases or DNA ligases is therefore eliminated by generating two PCR fragments with overlapping ends.

This method is illustrated in Figure 6.7. In separate PCRs two fragments of the target gene sequence are amplified (*am1* and *am2*). Each reaction uses one flanking primer that hybridises to DNA at one side of the target sequence (primer **a** or **d** in Figure 6.7) and one internal primer that hybridises to one side of the intergenic region to be removed (primer

b or **c** in Figure 6.7). Since the product generated in a PCR incorporates the primers, the wild type sequence will not be amplified. Primers **b** and **c** were designed such that their 3' ends hybridise to the template sequence on one side of the deletion, and their 5' ends are complementary to template sequence on the other side of the deletion. Because the two internal primers overlap, it should then be possible to fuse the two fragments AB and CD, generated in the first PCR, by denaturing and annealing them in a subsequent primer extension reaction. The overlap allows one strand from each fragment to act as a primer on the other, and extension of this overlap should result in the mutant fusion product in Figure 6.7. Even though the annealing of the short overlap between the two fragments may occur at low frequency, the inclusion of additional flanking primers (**a** and **d**) allows the fusion product that is formed to be amplified by PCR.

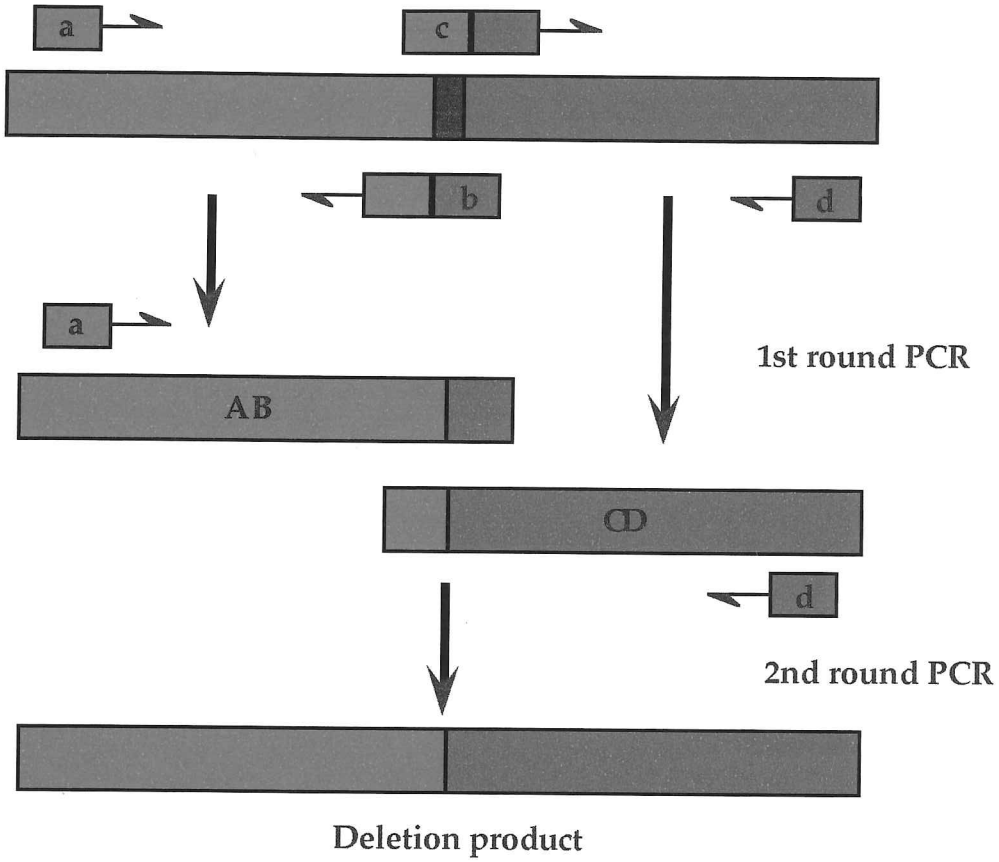
Primers **a** and **d** were designed so that they included *Sal*I restriction sites at their 5' ends to allow incorporation of the resulting fragment into the *Sal*I site of the pSVP27A polylinker. These two outside primers were also used to incorporate *Sal*I restriction sites on the ends of a non-mutated fragment of DNA containing *am1* and *am2* to facilitate cloning into the shuttle vector. Figure 6.7 shows the sequence of primers **a** and **d**. As well as the *Sal*I restriction site a short region of 'junk' DNA was also included at the 5' end to ensure efficient digestion with the *Sal*I restriction enzyme. Primers **b** and **c** were designed so that the resulting fusion product would contain *am1* and *am2* joined in the same open reading frame.

The length of each hybridising section of olionucleotide was crucial as this determines its melting temperature and therefore the temperature at which it will anneal to the template. The melting temperatures were roughly calculated using the following formula.

$$T_m\ (°C) = 4(G + C) + 2(A + T)$$

Primers **a** and **d** and the 3' regions of primers **b** and **c** were all designed to have similar melting temperatures (72°C according to the above formula). The annealing temperature of these regions was therefore calculated to be about 62°C. The 5' region of primers **b** and **c** were designed so that this portion of the olionucleotide that was to overlap with the other fragment in the fusion reaction had a calculated denaturation temperature that matched the annealing temperature to be used in the primary PCR reaction. Three PCR reactions were carried out using the following conditions.

Denaturation	**94°C**	**1 minute**
Annealing	**65°C**	**1 minute**
Extension	**72°C**	**2 minutes**



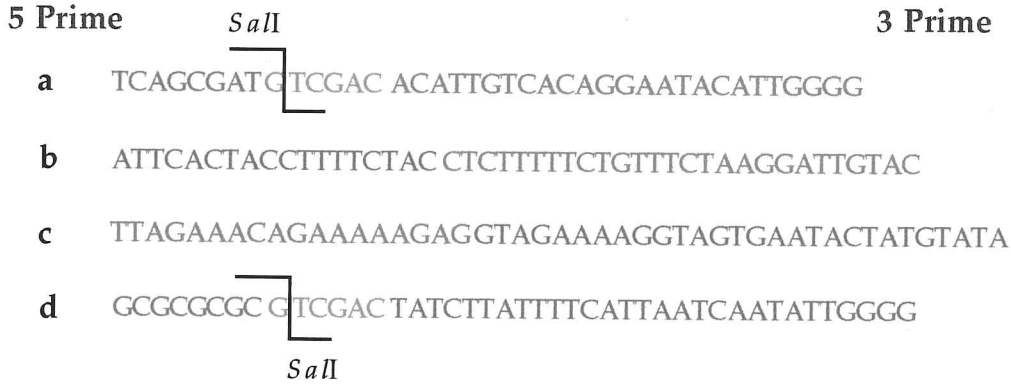


Figure 6.7 Schematic diagram showing how deletion mutations may be generated by the overlap extension method (adapted from Steffan *et al.*, 1989).
The sequence to be deleted is depicted by the black box. The red and purple boxes indicate the sequence on either side of the point of deletion (red harbours *am1* and purple *am2*). Primers '**b**' and '**c**' are designed such that their 3' ends hybridise to template sequence on one side of the deletion, and their 5' ends are complementary to the template sequence on the other side of the deletion. The **AB** and **CD** products generated from the PCR reaction using these primers are therefore overlapping at the deletion point.
The sequence of the primers **a-d** is shown, the colours indicate the regions of DNA sequence to which each section of the primer hybridises. The green sequence is a *SalI* site that was added to facilitate cloning of the resulting fragment in the shuttle vector. The blue sequence is 'junk' DNA that serves no purpose other than to ensure efficient digestion with *SalI*.

The template DNA (pSVP27AM1/2+) was added once 85°C had been reached to 'hot start' the reaction and therefore reduce the chance of false priming. 25 cycles were carried out with an extension time of 7 minutes in the last cycle to ensure the generation of full length product. Two primers were used in each reaction as follows:

	Primer 1	Primer 2	Expected size of product (kb)
Reaction 1	**a**	**d**	**4.00**
Reaction 2	**a**	**b**	**2.15**
Reaction 3	**c**	**d**	**2.05**

Reaction 1 was designed to generate a non-mutated *Sal*I fragment containing intact *am1* and *am2* genes for insertion into pSVP27A for subsequent investigation into the expression of the two genes. Reactions 2 and 3 were designed to generate fragments of DNA containing *am1* and *am2* genes respectively that could then be joined in a second round of PCR to form a continuous novel 130-kDa gene.

1/10 of the product of each reaction was run on an agarose gel to determine the success of the reaction (Figure 6.8a). As can be seen from the gel, reactions 2 and 3 produced a high concentration of product of the predicted size. Reaction 1 was less successful although a faint band was visible. A longer extension time (4 minutes) and 30 cycles was found to be necessary for reaction 1 to generate sufficient 4 kb product to clone into pSVP27A (Figure 6.8b). The cloning of this 4 kb fragment is described in section 6.3.2.

The products of reactions 2 and 3 were purified using a Gene Elute spin column as described in Section 2.2.5 and used in a second round of PCR using primers **a** and **d** to generate the mutated 4 kb fragment. After repeated attempts at a variety of annealing temperatures the best results were obtained at 40°C (Figure 6.8c). However, when the insert was eluted from the gel, digested with *Sal*I and used in a ligation with *Sal*I cut pSVP27A no transformants were obtained containing the insert. The problem was thought to be due to the low concentration of insert DNA and an attempt was made to increase the concentration of insert DNA be pooling DNA cut from several lanes of gel. It did however prove impossible to obtain DNA of high enough concentration. This may be because both the Geneclean II and Gene Elute spin column methods of DNA extraction tried here involve a precipitation step to supply the DNA in a small enough volume for use in the ligation which in itself involves the loss of DNA.

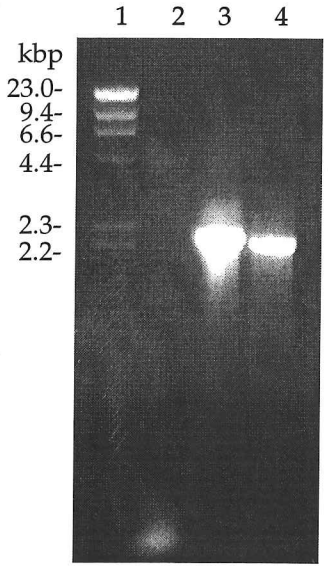


Figure 6.8a 0.8% agarose gel stained with ethidium bromide showing the products of the first round PCR reactions.
Reaction 1 (lane 2), reaction 2 (lane 3) and reaction 3 (lane 4). All three reactions had 25 cycles with an extension time of 2 minute. *Hind*III digested lambda DNA size markers were run in lane 1.

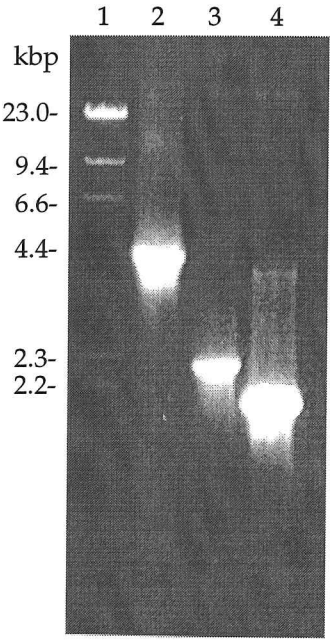


Figure 6.8b.
As for Figure 6.8a except that reaction 1 has been carried out for 30 cycles with the extension time increased to 4 minutes.

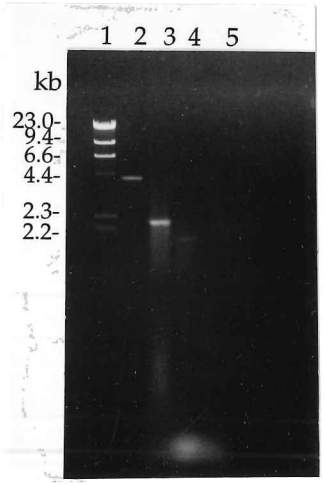


Figure 6.8c 0.8% agarose gel stained with ethidium bromide showing the product of the second round PCR reaction.
Second round PCR reaction (annealing temperature 40°C) with primers **a** + **d** in lane 2. Lanes 3 and 4 are positive controls using the same template but primers **a** + **b** and **c** + **d** respectively. Lane 5 is a negative control using primers **a** + **d** but no DNA.

A further attempt was made to increase the concentration of insert by using the 4 kb product of round 2 in a third round of PCR using primers **a** and **d**. This also proved unsuccessful. At this stage the time limitations of the project meant that no further attempts could be made to generate the chimeric AM1/AM2 protein.

6.4 Expression of *am1* and *am2* in Bt ssp. *israelensis* IPS78/11

The 4 kb fragment generated in the first round of PCR was digested with *Sal*I repurified from an agarose gel using a Gene Elute spin column and ligated into *Sal*I digested, dephosphorylated pSVP27A in an overnight reaction at 15°C. 1 µl of the ligation mix was transformed into electrocompetant JM109 cells and transformants selected for on LB plates containing 100 µg/ml ampicillin. The plasmid content of the colonies was analysed by restriction endonuclease digestion and agarose gel electrophoresis of small scale DNA preparations. One positive clone was obtained out of eight colonies, this was found by further restriction digest analysis to have *am1* and *am2* orientated in the clockwise direction placing these genes under the control of the *cytA* promoter, this clone was termed pSVP27AM1/2+. pSVP27AM1/2+ DNA was used to transform IPS78/11 and transformants were selected on LB containing 10 µg/ml chloramphenicol. The presence of the correct plasmid in these clones was verified by restriction analysis and hybridisation using DIG-labelled 17A.1 probe. This subcloning procedure is represented in Figure 6.9.

Colonies containing the pSVP27AM1/2+ plasmid were streaked onto CCY agar containing 10 µg/ml of chloramphenicol to induce expression of the *am1* and *am2* genes. On sporulation, observation by phase contrast microscopy identified the presence of very large, irregular, phase dark inclusion bodies that in some cases were much larger than the spore and filled the cell. Crystal preparations from this recombinant strain were again purified on a discontinuous, 1.97 M, 2.1 M, 2.3 M, sucrose gradient. Bands formed at both interfaces. Both bands were removed, washed several times with ice-cold water and the purity assessed by phase contrast microscopy. The band formed at the 1.97/2.1 M interface was found to be composed of a mixture of cell debris, crystals and spores, whereas the band at the 2.1/2.3 M interface was composed almost entirely of crystals. Pellet material was also examined and found to be contain mainly spores. The recovery of the purified inclusion preparation was calculated to be approximately 2 g per litre of culture.

Purified inclusions were then analysed by SDS-PAGE alongside 17A crystals (Figure 6.10). Two major bands were evident, one at about 70-kDa as expected and a second band that migrated at approximately 17-kDa.

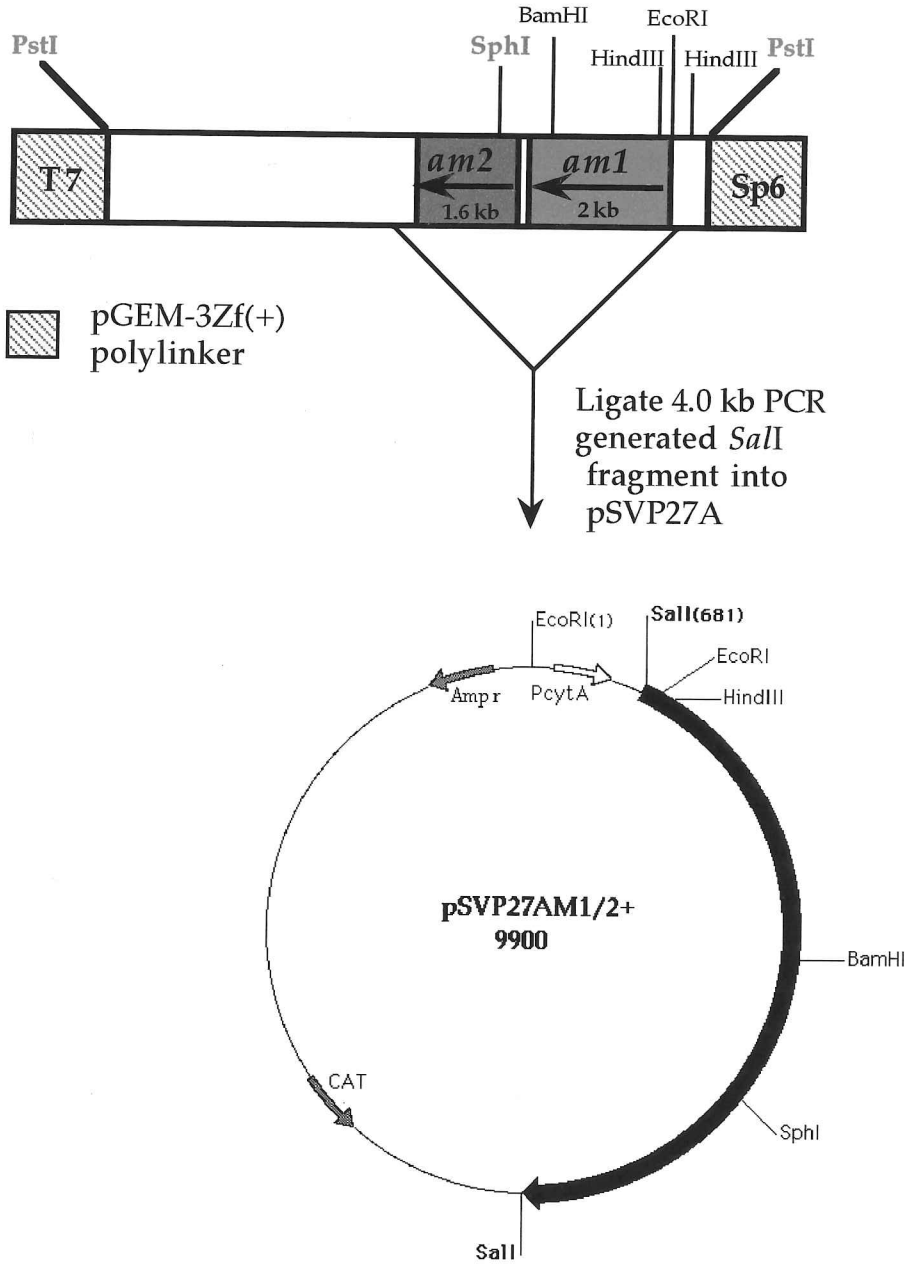


Figure 6.9 Subcloning of a 4.0 kb PCR generated *Sal*I fragment containing the complete open reading frames *am1* and *am2* into the Bt shuttle vector pSVP27A.
The construct pAM10 is shown as a bar map. The positions of the T7 and Sp6 priming sites are indicated. The position of *am1* is shown by a red box and that of *am2* by a purple box.
Construct pSVP27AM1/2 + is represented by a vector circle map. The thick black arrow indicates the direction of *am1* and *am2* transcription.

The identities of bands were investigated by blotting onto Problott membrane and N-terminal sequence analysis.

Only the upper of the two bands produced readable N-terminal sequence. Mixed sequence containing residues from the N-terminal sequence of both AM1 and AM2 was obtained for this band confirming that both AM1 and AM2 had been expressed from the clone pSVP27AM1/2+. Figure 6.11 shows a comparison between the N-terminal sequences of this band and those obtained from expression from pSVP27AM1+.

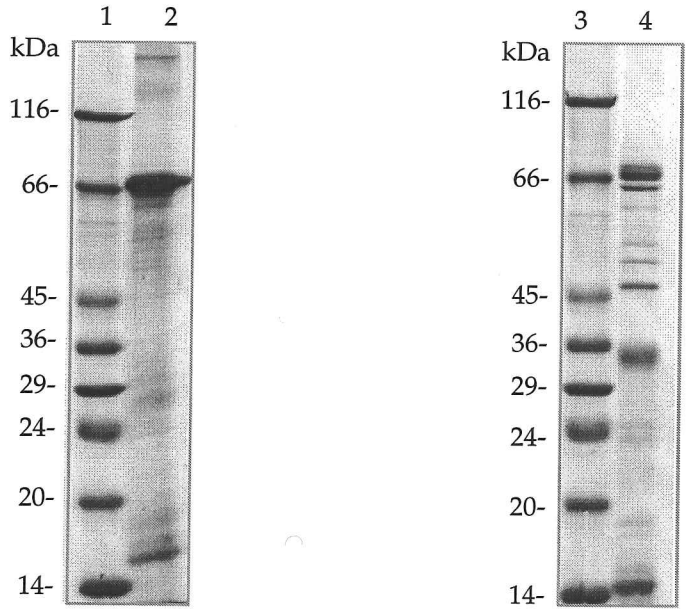


Figure 6.10 Coomassie blue stained 13% SDS-PAGE analysis of sucrose gradient purified inclusions from Bt IPS78/11 harbouring pSVP27AM1/2+ in panel A (lane 2) and Bt ssp. *fukuokaensis* 17A in panel B (lane 4) for comparison. Molecular weight markers are shown in lanes 1 and 3.

AM1+ (70-kDa)	**M**	**N**	**S**	**Y**	**Q**	**N**	**K**	**E**	**Y**	**E**	**I**	**L**	**D**	**A**	**S**	**R**	**N**	**N**	**S**
AM1+ (24-kDa)	**M**	**Y**	**T**	**N**	**T**	**M**	**K**	**N**	**K**	**L**	**K**	**I**	**E**	**T**	**T**	**D**	**Y**	**E**	**I**
AM1/2+ (70-kDa)	**M**	**N**	**T**	**Y**	**Q**	**N**	**K**	**N**											
		Y		**N**	**T**	**M**													

Figure 6.11 Comparison of the N-terminal sequence obtained from the proteins expressed from pSVP27AM1+ (AM1+) and pSVP27AM1/2+ (AM1/2+) in IPS78/11. The bottom line shows the mixed sequence results obtained from the 70-kDa band expressed from this construct. All of these residues can be found in the above two sequences indicating that this band is composed of two proteins.

6.5 Inclusion Solubilisation

Experiments were carried out to assess the solubility of purified inclusions from Bt IPS78/11 harbouring pSVP27AM1/2+ to determine if the solubility of AM1 had been affected by isolation from other 17A crystal components. This is important as Cry proteins need to be solubilised in the insect gut before they can exert a toxic effect. As can be seen from the SDS-PAGE gel in Figure 6.12, the 70-kDa band was never completely soluble in 50 mM buffer in the absence of DTT and even in its presence the pH had to be taken as high as 12.5 before complete solubility was achieved. This is markedly different from the solubility of the same protein in the native 17A crystal. As discussed in Chapter 3, solubilisation of the complete 17A inclusion was achieved at pH 10 and pH 11 with or without DTT respectively.

This marked change in the pH needed to solubilise the cloned protein relative to its native state in the 17A inclusion may reduce its toxicity to *Dacus oleus* as AM1 may no longer be soluble in the insect gut.

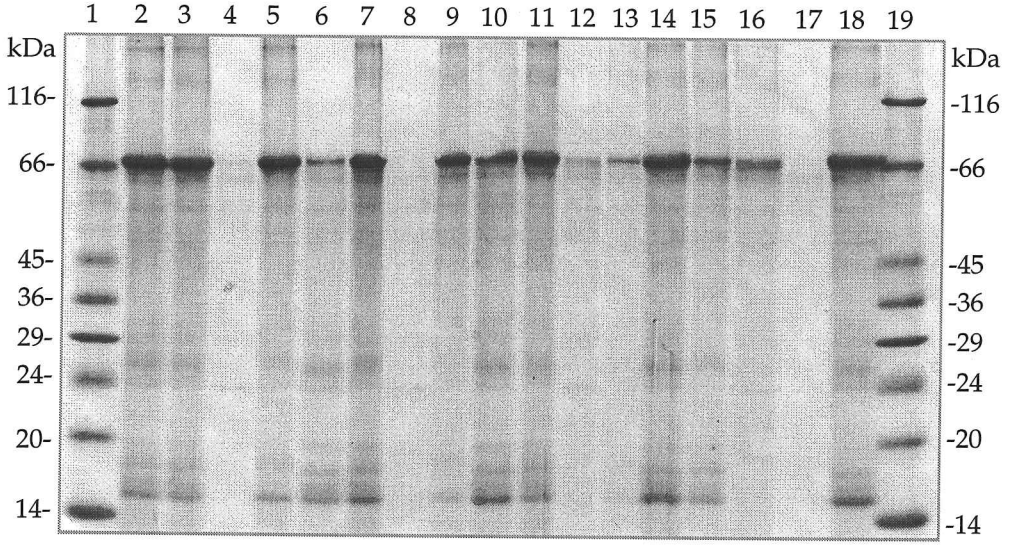


Figure 6.12 Coomassie blue stained 13% SDS-PAGE gel of 10 µg of AM1 protein inclusions purified from IPS78/11 (pSVP27AM1/2+) and solubilised in 50 mM Na_2CO_3 buffers of varying pH with or without the presence of 10 mM DTT (Photo no. 41E3).

Lane 1	Molecular weight marker
Lane 2	10 g purified AM1 inclusions
Lane 3&4	pH 11 -DTT, pellet and supernatant fractions
Lane 5&6	pH 11 +DTT, pellet and supernatant fractions
Lane 7&8	pH 11.5 -DTT, pellet and supernatant fractions
Lane 9&10	pH 11.5 +DTT, pellet and supernatant fractions
Lane 11&12	pH 12 -DTT, pellet and supernatant fractions
Lane 13&14	pH 12 +DTT, pellet and supernatant fractions
Lane 15&16	pH 12.5 -DTT, pellet and supernatant fractions
Lane 17&18	pH 12.5 +DTT, pellet and supernatant fractions
Lane 19	Molecular weight marker

6.6 *In vivo* toxicity assay

Initial studies into the entomopathogenic properties of Bt ssp. *fukuokaensis* 17A had reported spore-crystal mixtures to be toxic to both adult (15%) and larval (87%) *Dacus oleae* (Karamanlidou *et al.*, 1991). In an attempt to determine the extent to which the AM1 and AM2 proteins were responsible for this toxicity, spore-crystal mixtures from Bt IPS78/11 harbouring pSVP27AM1/2+ and wild type 17A were freeze-dried and sent for bioassay to Professor R. Prota at the at the Instituto di Entomologia Agraria (Italy). Bioassays were carried out on four to five day old larvae and the mortality recorded after 48 hours (Figure 6.13).

Spore-crystal mix (500 µg/ml)	Number of dead larvae in each replicate.			Total dead (out of 30)	Mortality (%)
Bt IPS78/11 (pSVP27AM1/2+)	2	5	5	12	40.0
wild-type 17A	3	6	7	16	53.3
Control	1	2	1	4	13.3

Figure 6.13 Bioassay of spore-crystal mix from Bt IPS78/11 harbouring the clone pSVP27AM1/2+ and wild-type Bt ssp. *fukuokaensis* 17A against 4-5 day old *Dacus oleae* larvae.
Each replicate contained 10 larvae. The control contained no spore-crystal mix.

Both inclusions containing the cloned AM1/AM2 proteins and 17A protein appear to be toxic to *Dacus oleae* larvae. This is little more than a preliminary assay, however it does appear that the AM1 and AM2 proteins are responsible for at least part of the toxicity of wild-type strain. The mortality rate for the control was low (13%) confirming that the bioassay protocol was satisfactory and that the observed results were due to the test samples.

6.7 Discussion

With the exceptions of Cry3A in Bt ssp. *tenebrionis* where *cry3A* transcripts and crystal protein antigens can be detected in vegetatively growing cells (Sekar, 1988) and Cry1I which is expressed from early stationary phase (Kostichka *et al.*, 1996) the expression of δ-endotoxin genes in Bt is generally restricted to the sporulation phase of the growth cycle

(Klier *et al.*, 1982; Wong *et al.*, 1983; Ward & Ellar, 1987). In this Chapter the expression of the putative crystal protein gene *am1* from Bt ssp. *fukuokaensis* 17A was studied.

Despite the fact that an apparently functional sporulation specific promoter region had been identified upstream of *am1* in clone pSVP27AM1- (see Chapter 5) crystal protein expression was not found to occur from this promoter in IPS78/11. This is perhaps not surprising as AM1 lacks the characteristic cysteine rich C-termini thought to be necessary for crystal formation (Höfte & Whiteley, 1989). As discussed in Chapter 5 the C-terminal region is encoded in a separate open reading frame (*am2*) immediately downstream of *am1* and this may be necessary for efficient expression of *am1*. Further work should be carried out to determine if *am1* can be expressed from its own promoter if the entire *am2* region is present.

When this same region of DNA was placed under the control of the *cyt1A* promoter, crystalline inclusions were formed and found to be composed of both AM1 and the first 24-kDa of the C-terminal region AM2 encoded by pSVP27AM1+. The expression of these proteins from the *cyt1A* promoter and not from the promoter upstream of *am1* may indicate that the *am1* promoter is non-functional. However, accumulation of these two proteins may also be due to their over expression under the highly efficient *cyt1A* promoter resulting in non-specific and non-native aggregation.

A second clone (pSVP27AM1/2+) was constructed so that the toxicity of inclusions containing AM1 and AM2 could be investigated. On the basis of the results obtained from a preliminary assay it appears that the AM1 and AM2 proteins could be responsible for 75% of the toxicity towards *Dacus oleae* larvae observed for the wild-type strain. Further experiments should be carried out to determine the relative contributions of AM1 and AM2 to this toxicity.

Attempts to generate a chimaeric AM1/AM2 gene by removal of the intergenic segment of DNA by overlap extension PCR were unsuccessful. In retrospect it may have been simpler to use primers **b** and **c** to generate the same restriction site on either side of the intergenic DNA. *am1* and *am2* could then have been joined by ligation. Construction of this clone from a pSVP27AM1/2- orientation would enable an interesting investigation into the expression and toxicity of the chimaeric product in relation to the expression and toxicity of AM1 and AM2 from the natural, dissociated gene arrangement. PCR could also be used to generate suitable sites in the intergenic region to facilitate the construction of two further clones containing isolated *am1* and *am2* genes This is discussed further in the final Chapter of this Thesis.

Seven

Chapter 7

Cloning of Novel 90-kDa Putative Toxin

7.1 Introduction

The previous Chapters describe the identification and cloning of a 70-kDa Cry10A type δ-endotoxin from Bt ssp. *fukuokaensis* 17A. In Chapter 3, a second Cry10A type protein was also identified in the related strain Bt ssp. *fukuokaensis* 84-I by N-terminal amino acid sequence analysis of a 90-kDa protein band. Bt ssp. *fukuokaensis* 84-I was first isolated and characterised in Japan by Yu *et al.* in 1991 and reported to have moderate levels of mosquitocidal activity (Yu *et al.*, 1991). The crystals were shown to be composed of a complex set of proteins similar to other diptera specific strains (Yu *et al.*, 1991) (Figure 3.4). Immunoblotting and PCR analysis implied that the proteins were likely to be novel (Chapter 3).

The similarity of the N-terminal sequences of the 70-kDa protein from 17A and the 90-kDa protein from 84-I raised the question as to how they were related throughout the length of the protein. To investigate this it was decided to clone, sequence and express both genes. Attempts could then be made to correlate differences in size and amino acid sequence with potential differences in stability, solubility and activity towards various insects. It was also hoped that it would be possible to conduct domain swapping experiments between the two proteins to further characterise the role of particular regions.

This Chapter describes the cloning of plasmid DNA fragments believed to harbour the gene for a novel Cry10A type 90-kDa protein from Bt ssp. *fukuokaensis* 84-I. Size enriched plasmid DNA libraries were constructed and screened using a digoxygenin (DIG)-labelled oligonucleotides based upon the N-terminal sequence of the 90-kDa protein. Construction of the plasmid DNA library and identification of positive clones was in principle the same as is described in Chapters 4 and 5.

7.2 Design and Production of Oligonucleotide Probes

In an attempt to clone the gene encoding the 90-kDa protein from Bt ssp. *fukuokaensis* 84-I, a synthetic oligonucleotide probe was designed based on the N-terminal amino acid sequence determined in Chapter 3 after consideration of Bt δ-endotoxin usage (Figure 4.3). Figure 7.1 shows the design of the 84-I oligonucleotide with respect to the N-terminal sequence of the 90-kDa protein.

The single stranded oligonucleotide 84-I.1 was labelled by incorporation of digoxygenin using terminal transferase as described in Chapter 4. The success of the labelling reaction was confirmed by direct detection of the tailed oligonucleotide probes in a dot blot experiment shown in Figure 7.2 (Section 2.2.9). Comparison of a serial dilution of 84-I.1 with that of a pre-labelled control enabled the yield of DIG-labelled oligonucleotide to be estimated at approximately 1 pmol/µl (Figure 7.2).

Oligonucleotide 84-I.1

```
5'                                                    3'
 M    N    P    Y    Q    N    N    D    E
ATG  AAT  CCA  TAT  CAA  AAT  AAT  GAT  GAA
```

Figure 7.1 The N-terminal sequence of the 90-kDa protein from Bt ssp. *fukuokaensis* 84-I and the corresponding 24 base oligonucleotide (84-I.1) selected for screening of transformants and designed by deduction from the codon usage table (Figure 4.3).

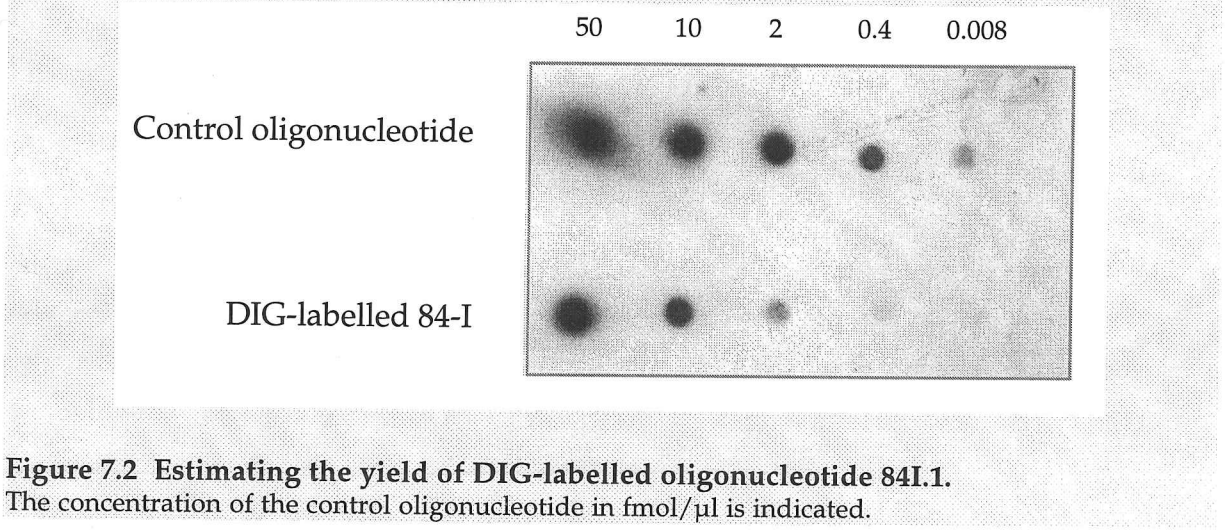


Figure 7.2 Estimating the yield of DIG-labelled oligonucleotide 84I.1.
The concentration of the control oligonucleotide in fmol/µl is indicated.

7.3 Results and Discussion

7.3.1 Location of the 90-kDa protein gene in Bt ssp. *fukuokaensis* 84-I

As described in Chapter 4, Bt δ-endotoxins are generally found on extra-chromosomal plasmids (Gonzalez *et al.*, 1981; Carlton & Gonzalez, 1985). Bt ssp. *fukuokaensis* 84-I has been reported to contain a single large 130 MDa plasmid (Yu *et al.*, 1991). Loss of this plasmid has been correlated with the formation of an acrystalliferous mutant (Yu *et al.*, 1991), indicating that the crystal proteins of 84-I are also plasmid encoded. Plasmid curing at 42°C was attempted by Dunn (1996) to confirm the results presented in Yu *et al.* (1991). Bt ssp. *fukuokaensis* 84-I was streaked onto nutrient agar and also CCY medium and incubated at 42°C for two days. Eighty colonies with unusual growth morphology were picked and streaked to single colonies, grown to sporulation at 30°C and analysed by phase-contrast microscopy. All the colonies examined appeared to retain crystal formation. Various other protocols were also employed to generate plasmid deficient mutants including the addition of novobiocin (2-5 mg/ml) and SDS (0.002% (w/v)) in the Nutrient Agar (Gonzalez *et al.*, 1981). These attempts also proved unsuccessful leading to the conclusion (contrary to the results of Yu *et al.*, 1991) that either the 130 MDa plasmid of Bt ssp. *fukuokaensis* 84-I is very stable, or that the crystal producing genes are chromosomally encoded.

Dunn then went on to clone the 24-kDa crystal component of 84-I from a preparation of plasmid DNA, indicating that the crystal genes were indeed plasmid encoded. However, it should be noted that the plasmid preparation from which the 24-kDa crystal gene was cloned also contained contaminating chromosomal DNA (as was found to be the case with the preparations from 17A described in Chapter 4), thus it is still possible that the gene is chromosomally encoded.

Bt δ-endotoxin genes have only very rarely been found to be chromosomally encoded (Klier *et al.*, 1982; Kronstad *et al.*, 1983). Therefore, in the absence of any firm evidence to indicate that the δ-endotoxin genes of 84-I are other than on a plasmid it was decided to attempt to clone the 90-kDa crystal gene from Bt ssp. *fukuokaensis* 84-I plasmid DNA. Large scale plasmid preparations were carried out as described in Chapter 4. 80-100 μg of DNA being obtained from each litre of culture.

About 5 μg of uncut plasmid DNA was analysed by agarose gel electrophoresis to check the plasmid profile. At least two bands were identified corresponding to large molecular weight plasmids (Figure 7.3). This is contrary to the findings of Yu *et al.*, (1991) where only one 130 MDa plasmid was identified. The strain from which this plasmid DNA was obtained was confirmed as 84-I by SDS-PAGE analysis of the crystal protein from a sporulating culture (see Figure 3.3 for the polypeptide profile).

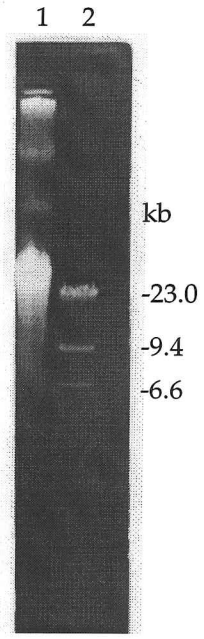


Figure 7.3 Plasmid profile of Bt ssp. *fukuokaensis* 84-I.
The DNA molecular weight marker is *Hin*dIII digested lambda DNA.
Lane 1 uncut 84-I plasmid DNA.
Lane 2 molecular weight marker.

7.3.2 Bt ssp. *fukuokaensis* 84-I DNA Hybridisation Studies and Plasmid Library Construction

Caesium chloride-gradient purified Bt ssp. *fukuokaensis* 84-I plasmid DNA was digested with several restriction enzymes (5 µg per reaction), electrophoresed on a 0.8% agarose gel and Southern blot analysed with the DIG-labelled 84-I.1 probe. Initially, the stringent wash was carried out at 0.1x SSC and the hybridisation temperature calculated according to the equation in Figure 4.4. However, as was found to be the case with 17A DNA and probes 17A.1 and 17A.2 no positive bands were identified. An optimisation procedure was carried out to identify the most suitable conditions for hybridisation (as described in Chapter 4) which were eventually determined to be 45°C with 0.5x SSC strigency wash.

Under these conditions, the oligonucleotide probe 84-I.1 was found to hybridise with *Hind*III, *Pst*I, and *Sph*I fragments of approximately 4.2 kb, 9.4 kb and 8.0 kb respectively. Two *Eco*RI bands of 6.0 kb and 2.1 kb were evident with the 2.1 kb band having by far the weaker signal (Figure 7.4).

In order to clone the 8.0 kb *Sph*I fragment, caesium-chloride gradient purified Bt ssp. *fukuokaensis* 84-I plasmid DNA (100 µg) was digested with *Sph*I in a total volume of 200 µl for eight hours. The digested plasmid was then electrophoresed on a 0.5% agarose gel in 1x TAE at 40 Volts for eighteen hours. Five bands from the plasmid digest between the 6.6 kb and 9.4 kb markers were excised, the DNA from each band repurified using a Geneclean II kit and resuspended in a volume of 20 µl. It was decided to carry out "blind" ligations with DNA from all five gel slices as had proved successful in the cloning of the 17A crystal protein gene (described in Chapter 4).

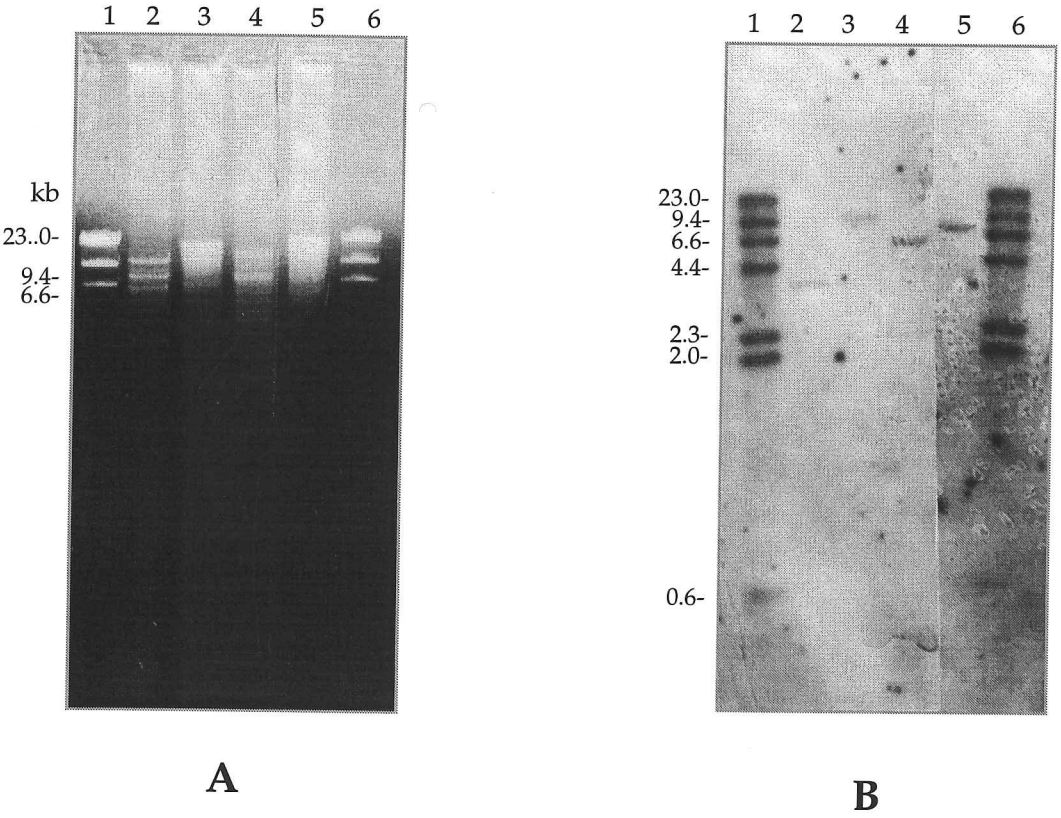


Figure 7.4 Southern blot analysis of Bt. ssp *fukuokaensis* 84-I plasmid DNA (photo no. 39E4).
Panel A: 0.8% agarose gel stained with ethidium bromide showing Bt ssp. *fukuokaensis* 84-I plasmid DNA digested with *Hin*dIII (lane 2), *Pst*I (lane 3), *Eco*RI (lane 4) and *Sph*I (lane 5). Digoxygenin-labelled *Hin*dIII digested lambda DNA size markers were run in lanes 1 and 6.
Panel B: Southern blot and hybridisation of Panel A with digoxygenin-labelled 84-I.1.

The total yield from each insert preparation was ligated to about 15 ng of *Sph*I-linearised, dephosphorylated pGEM-3Zf(+) (approximately a 1:3 molar ratio of insert to vector) at 16°C in a total volume of 20 µl for 16 hours. The ligase was then heat inactivated and 1 µl of each reaction mixture transformed into JM109 as described in Section 4.4. About 50 colonies were obtained per transformation of which about 40% were white. Bulk preparations of several white colonies were carried out as detailed in Section 4.4 and the DNA subjected to Southern blot hybridisation to identify one preparation containing a possible clone. A second round of hybridisation with individual small scale plasmid preparations identified a single preparation with the correct size hybridising band (pAM84.8), this was used for further investigation.

pAM84.8 DNA was prepared using the Perfectprep™ plasmid DNA kit and 1.2 kb of unambiguous sequence data was obtained for each clone (0.6 kb from either end of the

insert) using the T7 and Sp6 universal primers as described in Chapter 4. A GCG BLAST search was run on the DNA sequence obtained but no significant similarity was found to any previously identified δ-endotoxin genes. Neither was any sequence found that would be likely to bind the 84-I.1 probe. However, about 6.8 kb of DNA still remained unsequenced, sufficient to contain the whole of a gene encoding a 90-kDa protein. The sequence was also found to contain a low G+C ratio compatible with that of Bt DNA.

Restriction mapping was carried out on pAM84.8 using enzymes known to cleave the pGEM-3Zf(+) polylinker (Figure 7.5). *Hin*dIII digestion of pAM84.8 generated three restriction fragments of sizes 5.1 kb, 3.5 kb and 2.4 kb. Subsequent blotting and hybridisation revealed that the 3.5 kb fragment harboured a sequence to which the DIG-labelled oligonucleotide hybridised (Figure 7.5). A full restriction map of pAM84.8 is given in Figure 7.6.

The 3.5 kb *Hin*dIII fragment was selected for subcloning in order to facilitate sequence analysis of the region that had bound the 84-I.1 oligonucleotide probe and thereby hopefully identify the gene encoding the 90-kDa crystal protein. It was also decided to subclone the 5.1 kb and 2.2 kb *Hin*dIII fragments so that they were readily available for sequence analysis if the 3.5 kb fragment was found to contain only part of the δ–endotoxin gene. In order to subclone the three *Hin*dIII fragments, about 50 ng of *Hin*dIII digested pAM84.8 was run on a 0.5% agarose gel in 1x TAE buffer. The three bands were excised, individually repurified using a Geneclean II kit and resuspended in a volume of 20 µl. Three ligation reactions were set up. The 3.5 and 2.2 kb insert fragments were ligated to *Hin*dIII linearised, dephosphorylated pGEM-3Zf(+) in an approximate 1:3 molar ratio of vector to insert. The 5.1 kb fragment already contained linearised pGEM vector so this was simply recircularised to form a subclone containing 1.9 kb of cloned DNA. After heat inactivation of the ligase 1 µl of each reaction was transformed into 50 µl of *E. coli* JM109 under the conditions described for cloning of the fragments in Chapter 4. Small scale plasmid preparations of transformant colonies were carried out to isolate DNA for all three subclones.

*Hin*dIII restriction digests were carried out to identify transformants that contained the correct sized fragments (Figure 7.7) and the identity of the 3.5 kb subclone (pAMscX) was confirmed by Southern blotting and positive hybridisation with 84-I.1 (Figure 7.7).

Sequencing reactions carried out on Perfectprep™ prepared pAMscX DNA from each end of the subclones using the T7 and Sp6 primers did not reveal sequence with significant similarity to any previously identified δ-endotoxin genes. Nor was any sequence identified to which the probe 84-I.1 could obviously bind. The whole of pAMscX was sequenced by 'gene walking', using oligonucleotides designed to the ends of previously

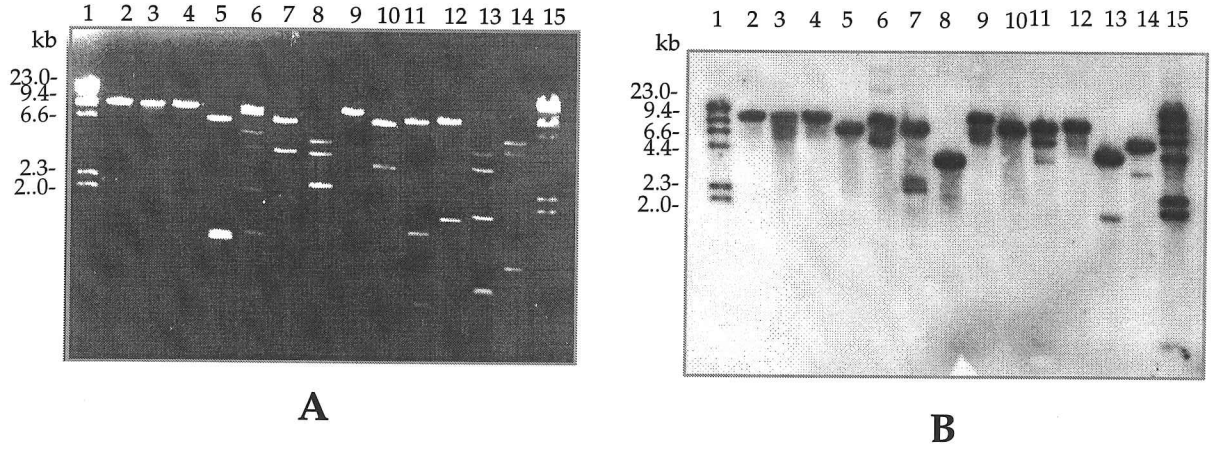


Figure 7.5 Restriction endonuclease analysis of clone pAM84.8. (photo no. 30E1)
Panel A: 0.8% agarose gel stained with ethidium bromide showing construct pAM84.8 digested with the following enzymes. The size marker in lanes 1 and 15 is *Hind*III digested Digoxygenin-labelled lambda DNA.

Lane 1 Size marker	Lane 9 *Sal*I
Lane 2 *Sac*I	Lane 10 *Sph*I
Lane 3 *Kpn*I	Lane 11 *Eco*RI
Lane 4 *Xba*I	Lane 12 *Bam*HI
Lane 5 *Acc*I	Lane 13 *Sca*I
Lane 6 *Hinc*II	Lane 14 *Nde*I
Lane 7 *Pst*I	Lane 15 Size marker
Lane 8 *Hind*III	

Panel B: Southern blot and hybridisation of Panel A with digoxygenin-labelled 84-I.1.

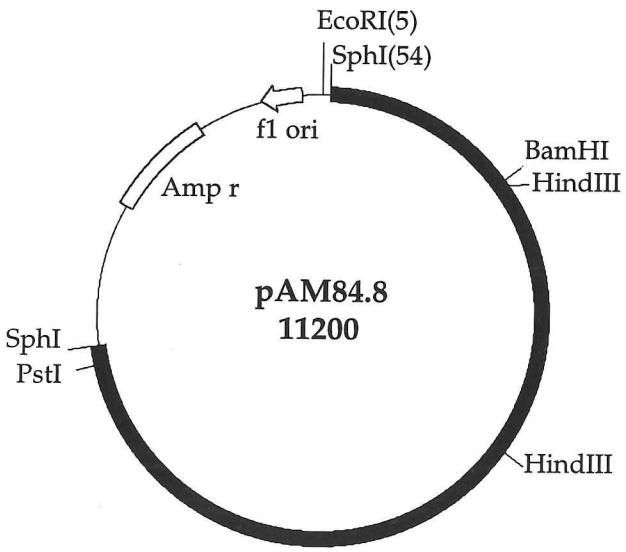


Figure 7.6 Restriction map of plasmid construct pAM84.8.
The 8.0 kb fragment (black box) is cloned into the *Sph*I (position 54) site of the vector pGEM-3Zf(+).
The ampicillin-resistance gene is represented as Amp r and the f1 origin is also indicated. Position 5 marks the start of the pGEM-3Zf(+) polylinker.

obtained sequence as primers for the next round of sequencing. The resulting sequences were compiled using the 'bestfit' program in the GCG package to generate 3.5 kb of unambiguous sequence data (not shown). BLAST searches carried out on the DNA sequence did not reveal any significant similarity to any previously identified δ-endotoxin genes, nor any other reported gene sequences. The sequence was also searched for a region showing obvious homology to the probe 84-I.1 but none was found. It therefore appears that despite the strong hybridisation signals of construct pAM84.8 and subclone pAMscX, it was nevertheless a false positive signal probably generated by region of DNA showing only moderate homology to the probe 84-I.1. The existence a such a region was indicated by the presence of a second weaker signal in the *Eco*RI digest in the original hybridisation shown in Figure 7.4.

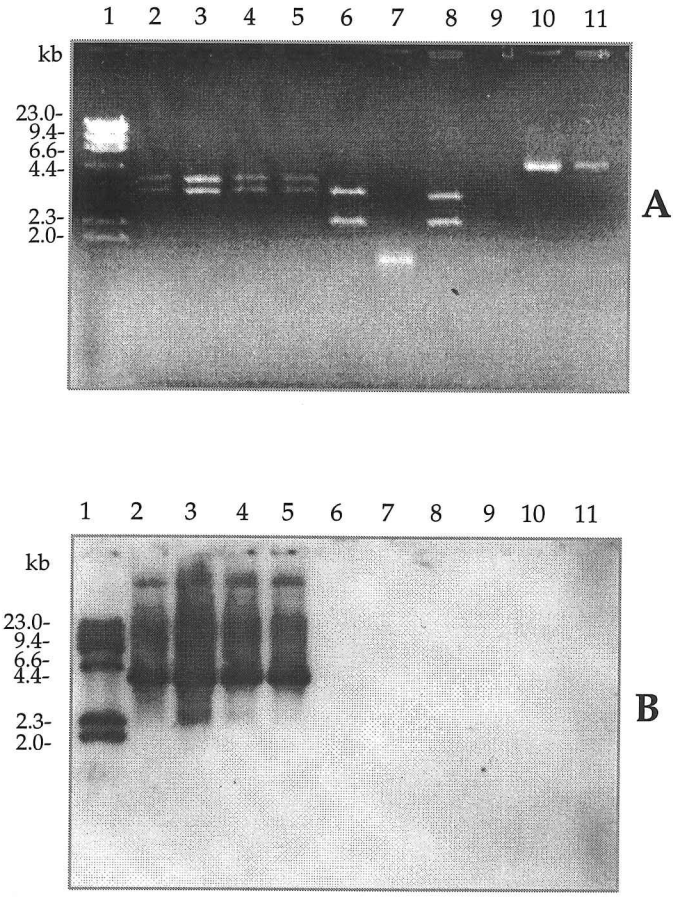


Figure 7.7
Restriction endonuclease analysis of the three *Hin*dIII subclones of pAM84.8 (photo no. 30E2).
Panel A: 0.8% agarose gel stained with ethidium bromide showing DNA from transformant colonies digested with *Hin*dIII. Lanes 2-5 contain subclone pAMscX DNA (3.5 kb insert fragment and 3.2 kb pGEM vector); lanes 6 and 8 contain successful subclones of the 2.2 kb *Hin*dIII fragment (lanes 7 and 9 contain DNA from unsuccessful transformants); lanes 10 and 11 contain successful subclones of the 5.1 kb fragment (linearised with *Hin*dIII). Digoxygenin-labelled λ-*Hin*dIII DNA size markers were run in lane 1
Panel B: Southern blot and hybridisation of Panel A with digoxygenin labelled 84-I.1.

It was decided that a second attempt would be made to clone the 90-kDa endotoxin gene from 84-I. Caesium chloride-gradient purified Bt ssp. *fukuokaensis* 84-I plasmid DNA was digested with several restriction enzymes and Southern blot analysed with 84-I.1 as previously described (Figure 7.8). In this blot the presence of a second weaker hybridisation signal is obvious in all four digests. For the *Sph*I digest the second signal is for a restriction fragment only slightly smaller than the main hybridising signal. This therefore explains how a hybridising *Sph*I fragment was cloned that appeared to be the correct size on restriction digest analysis but did not contain the 90-kDa δ-endotoxin gene. The difference between the weak signal seen in Figure 7.8 and the strong hybridisation signal seen in Figures 7.6 and 7.7 may be explained by the relative number of copies of the hybridisation region available in the total DNA digest in Figure 7.8 compared to the cloned DNA in Figure 7.6 and 7.7.

Attempts were made to clone the 10 kb hybridising *Pst*I fragment (Figure 7.8) but before this could be achieved information was received that the 90-kDa protein gene had been cloned by another group working on the same strain (Lee & Gill, unpublished). This gene has since been named Cry20A. At this point it was already obvious that a δ-endotoxin gene had been successfully cloned from Bt ssp. *fukuokaensis* 17A (Chapter 4). It was therefore decided that it would be more worthwhile to pursue the sequence analysis and characterisation of this novel protein, than spend time cloning a gene already in the process of being characterised. Attempts to clone the 90-kDa protein of Bt ssp. *fukuokaensis* 84-I were therefore abandoned at this stage.

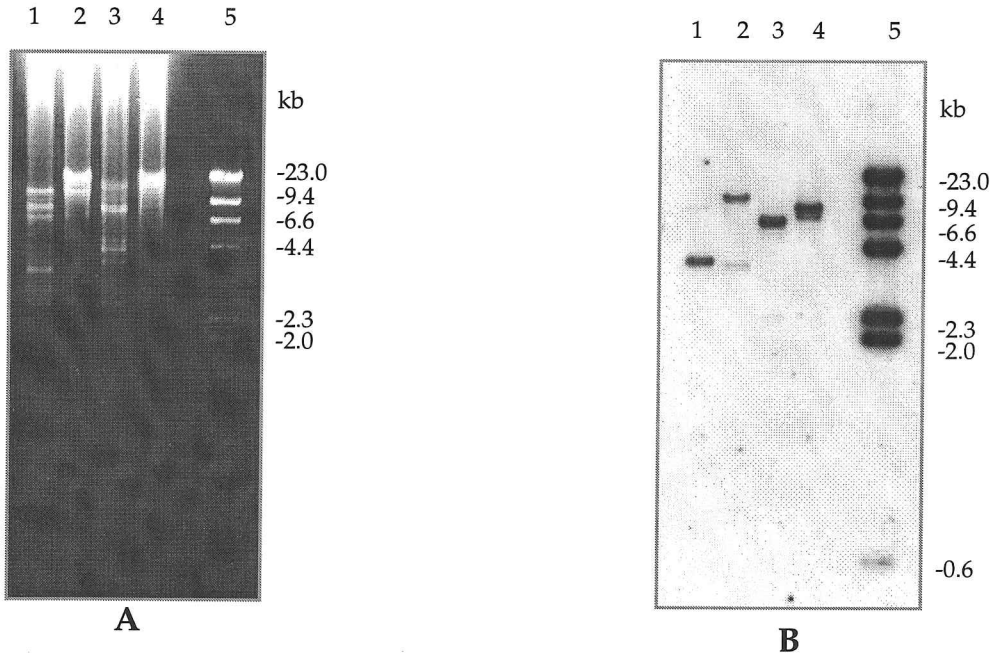


Figure 7.8 Southern blot analysis of Bt. ssp *fukuokaensis* 84-I plasmid DNA.
Panel A: 0.8% agarose gel stained with ethidium bromide showing Bt ssp. *fukuokaensis* 84-I plasmid DNA digested with *Hin*dIII (lane 1), *Pst*I (lane 2), *Eco*RI (lane 3) and *Sph*I (lane 4). Digoxygenin-labelled λ-*Hin*dIII DNA size markers were run in lane 6.
Panel B: Southern blot and hybridisation of Panel A with digoxygenin-labelled 84-I.1 (photo no. 908D1).

7.3.3 Phylogenetic Analysis of Cry20A from Bt ssp. *fukuokaensis* 84-I

In this section it should be noted that Orf1 and Orf2 refer to proteins of about 20-kDa not to the OFR1 and ORF2 regions (corresponding to the N and C-terminal regions) of the two gene arrangement of Cry10A that were discussed in previous Chapters. This unfortunately confusing nomenclature was used in previous Chapters so that direct reference could be made to the original and subsequent papers which characterise these proteins. This nomenclature will continue to be used throughout the rest of this work.

The sequence of Cry20A is, as yet, unpublished but is available from the EBI database (reference U82518). The N-terminal amino acid sequence of Cry20A (as deduced from the DNA sequence) was found to be identical to that determined for the 90-kDa protein of 84-I in Chapter 3 (see Figure 3.9). The *cry20A* gene encodes a 753 amino acid protein of 86,147 Da. It therefore seems likely that Cry20A undergoes little if any proteolysis in the native strain. Cry20A therefore differs from *am1* and *cry10A* (proteins with similar N-termini), which also encode proteins of about 80-kDa but are C-terminally processed to about 70-kDa. When attempts were made to express the cloned *cry20A* gene, expression was found to be poor, no stable product was obtained and the protein could not be subjected to SDS-PAGE analysis (Gill, personal communication).

Sequence analysis downstream of *cry20A* did not identify any ORF2 region such as had been identified downstream of *am1* and *cry10A* (Gill, personal communication). These initial observations indicate that Cry20A may be more closely related to Cry11A that to the Cry10A-type proteins. Cry11A also has a very similar N-terminus (see Figure 3.10) to the above proteins, has no ORF2 downstream region and shows little C-terminal proteolysis in the native strain (the gene encodes a 72,356 Da protein which is expressed at about 70-kDa.

Although it is not believed to be required for high level expression in Bt (Chang *et* al., 1993; Dervyn *et al.*, 1995), a 20-kDa protein (P20) has been shown to increase Cry11A production in *E. coli* by a post-translational stabilisation (Visick & Whiteley, 1991). P20 may therefore act as a chaperone protein to initiate, facilitate, or stabilise crystal formation. *cry11A* and *p20* are organised into a single transcriptional unit along with a third protein gene *p19*. *p19* shows significant similarity to the *orf1* genes of the two Bt ssp. *kurstaki cry2A* and *cry2C* operons. In addition, the Cry11A protein has regions which show similarity to both Cry2A and Cry2C toxins (Kalman *et al.*, 1993).

The role of P19 remains to be determined. It was proposed that P19 may have a role in Cry11A synthesis in Bti, but no obvious requirement has been found (Dervyn *et al.*, 1995). P19 may still however prove to be a chaperone protein. Firstly, the *orf1* genes show good sequence conservation in the operons described and secondly, P19 amino acid composition is unusual in that 11.7% of the amino acids are cysteine residues (Dervyn *et al.*, 1995). The C-terminal halves of the 130-kDa δ-endotoxins are also rich in cysteine residues and are involved in crystallisation of the polypeptide chains. Since Cry11A and Cry2 proteins do

not have this C-terminal feature it has been suggested (Dervyn *et al.*, 1995) that the P19 protein plays a role in protein-protein interactions either by conferring a particular crystal structure or allowing the co-assembly of the Cry11A inclusion. In this case, the effect of the P19 deletion would only be seen in the native strain.

As previously described in Chapter 6, a similar *orf1* protein of 24-kDa has been described in 84-I (Dunn, 1996). It was therefore hypothesised in Chapter 3, that the presence of Orf1 in 84-I, may be connected to the presence of what appeared from its size and N-terminal sequence to be a C-terminally truncated Cry4-type protein (later identified as Cry20A). The discovery that cloned and expressed Cry20A protein is unstable lends further support to this theory. The cloned genes for *cry20A* and *orf1* are currently in different laboratories. It would be interesting to investigate co-expression of these two genes.

In order to further characterise the *cry20A* gene and determine if it was indeed more closely related to *cry11A* than to *am1* (from the related strain 17A), amino acid sequence alignments were carried out using a Clustal 250 weight table as described in Chapter 5. Since the initial analysis of *am1* was carried out, a *cry10A*-type protein gene has been identified and cloned from the strain Bt ssp. *jegathesan* (Rosso & Delécluse, 1997). The similarity between this gene arrangement and those of *am1* and *cry10A* was immediately obvious and it was therefore decided to include this gene in the subsequent analysis. The sequence has not been yet been published but is available on the EBI data base (Y07603).

The amino acid sequence of Cry20A was therefore aligned with that from Cry4A, AM1 and Cry19A. Only the ORF1 regions of Cry10A, AM1 and Cry19A were used and the corresponding N-terminal regions of Cry4A and B (see Chapter 5 for details). The results of this alignment are shown in the table and phylogenetic tree in Figures 7.9a and 7.9b respectively. AM1 and Cry19A appear to be the most closely related sequences of any discussed in this section. This is particularly interesting in the light of the initial observation in Chapter 3 as to the similarity of the polypeptide profiles of the two host strains (Bt ssp. *fukuokaensis* and Bt ssp. *jegathesan* respectively). For ease of reference, proteins will be described as either short (70 to 90-kDa) or long (approximately 130-kDa). Based on the sequence alignment, all seven proteins appear to have evolved from a common ancestor to form three distinct groups which mirror the structural organisation of the genes. What is immediately obvious is that Cry20A is not (as was hypothesised above) a particularly close relative of Cry11A, but shows greater sequence similarity to the five proteins with associated ORF1 and ORF2 regions. The similarity appears to be greatest to those proteins with separate ORF2 regions.

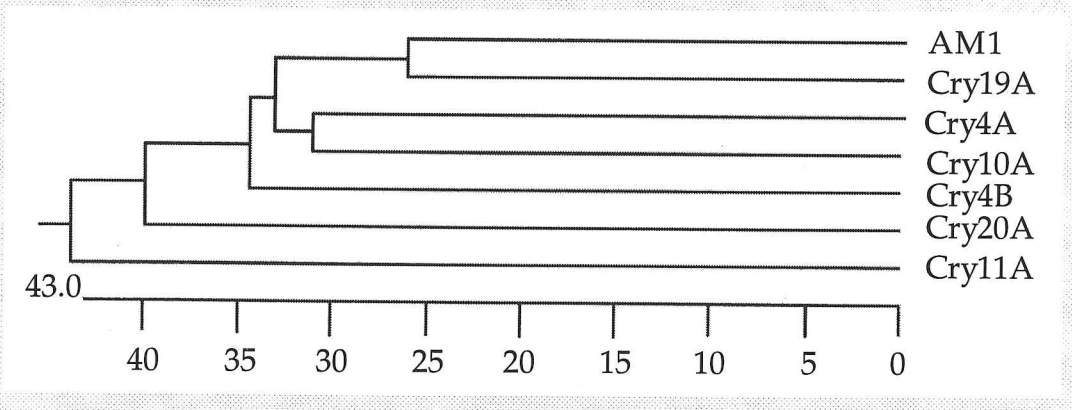
Group 1 is composed of Cry11A which shows the least sequence similarity to any of the other proteins. Cry11A is a short protein with no ORF2 region that is associated with

Percentage Similarity

	Cry20	AM1	Cry19A	Cry4A	Cry4B	Cry10A	Cry11A	
Cry20		20.2	20.1	17.8	18.5	21.5	14.0	Cry20
AM1	76.4		48.9	34.5	32.3	34.4	12.8	AM1
Cry19A	77.9	55.3		34.7	33.3	35.7	14.6	Cry19A
Cry4A	80.5	65.3	68.0		31.8	40.9	13.0	Cry4A
Cry4B	79.6	67.9	67.8	68.3		32.7	13.5	Cry4B
Cry10A	80.5	66.9	67.1	63.3	68.2		14.6	Cry10A
Cry11A	87.1	86.0	85.0	85.6	85.4	84.2		Cry11A
	Cry20	AM1	Cry19A	Cry4A	Cry4B	Cry10A	Cry11A	

Percentage Divergence

Figure 7.9a Protein sequence relationships using a PAM250 residue weight table.
Only the N-terminal regions of Cry4A and B that correspond to the ORF1 regions of AM1, Cry10A and Cry19 are used in this alignment.
The table represents the initial matrix used in the first process of the multiple alignment. The percentage similarity between individual pairs can be determined from the upper right area in the table. Percentage divergence between sequences is shown in the lower left portion of the table and is calculated by comparing sequence pairs in relation to the phylogenetic tree (Figure 7.9b).



7.9b Phylogenetic tree based on the sequence relationships described in Figure 7.9a.
The length of each pair of branches represents the distance between sequence pairs. The scale beneath the tree measures the distance between sequences; units indicate the number of substitution events.

an *orf1* type protein (Dervyn *et al.*, 1995).

Group 2 is composed of Cry20A, a short protein with no OFR2 region. Cry20A does not appear to be encoded in an operon with an Orf1 protein although the host strain (84-I) does express an Orf1 protein from another genetic locus (Dunn, 1996).

Group 3 is composed of both short: AM1, Cry19A and Cry10A and long: Cry4A and Cry4B proteins. All these proteins have an ORF1 and an ORF2 region. In the short proteins they are expressed separately and in the long proteins as a single polypeptide chain, but in both cases the ORF2 region appears to be essential for ORF1 synthesis (Rosso & Delécluse, 1997). The ORF2 region can therefore be considered as a chaperone. Group 3 proteins are not encoded in operons with Orf1 proteins and as yet, no Orf1 proteins have been identified in the host strains. From these groupings, and the ancestral distances implied between them, clues are beginning to emerge as to how the Group 3 δ-endotoxins may have evolved. This may also shed some light on the question raised in Chapter 5 as to whether the two-gene arrangements of AM1 and Cry10A (and now Cry19A) are the progeny or the progenitors of the 130-kDa Cry4A and B proteins.

The nature of this process of evolution was suggested by Dervyn *et al.*, 1995 before many of the above sequences were available for comparison. The old nomenclature was used in this hypothesis. They proposed that, on the bases of similar gene organisation and sequence similarity (22.8% according to Dervyn *et al.*, 1995) the CryIVD (Cry11A) polypeptide could be classified as a CryII (Cry2) rather than a CryIV (Cry4) protein. On this basis, evolution from a common ancestor operon encoding three open reading frames (*orf1, orf2* and a *cry* gene) could be proposed involving the following steps; (i) loss of *orf1* and *orf2*, which led to the monocistronic organisation of the *cryI, cryIII* (Cry3) and *cryIV* gene families and (ii) loss of the *orf2* gene, leading to the *cryIVD* operon, in contrast to the *cryII* family which has conserved *orf2*. According to this hypothesis Dervyn *et al.*, (1995) also proposed that the function of the Orf1 and Orf2 proteins, if necessary, would have to be compensated for by other proteins.

Alignment of Cry2A with the above seven proteins using a PAM250 weight table did indeed show that Cry11A was more closely related to Cry2A than to any of the other proteins (15.2% similarity, the discrepancy in this value and that obtained by Dervyn *et al.* (1995) is due to the different methods used). Based on the above hypothesis and the new data discussed in this and previous Chapters, I propose that a common ancestor operon encoding three open reading frames (*orf1, orf2* and a short *cry* gene) could have evolved in the following steps; (i) loss of the *orf2* gene leading to the *cry11A* operon, in contrast to the *cry2* family which has conserved *orf2*; (ii) movement of the *orf1* gene out of the operon to a new genetic locus leading to *cry20A*; (iii) insertion of the gene for a 60-kDa ORF2 protein immediately downstream of the *cry* gene leading to the two-gene arrangements of AM1, Cry10A and Cry19A (the ORF2 gene may or may not already have been present at another

genetic locus in the same strain and expressed into the crystal); (iv) deletion of the small intergenic fragment of DNA between the *cry* gene and the gene for the ORF2 protein leading to the single-gene arrangements of *cry4A* and *cry4B* and the subsequent expression of a 130-kDa protein.

Evidence for step (ii) may be found in the DNA surrounding the *orf1* gene. Dunn (1996) identified several open reading frames with significant similarity to previously identified transposable elements surrounding the cloned *orf1* gene from Bt ssp. *fukuokaensis* 84-I. An IS*240* like insertion sequence is located immediately upstream of *orf1*, whilst further upstream, an IS*150* like pair of open reading frames and the gene for a site-specific recombinase shown to be involved in Class II transposition formed what was possibly part of a novel Class II transposon. Such transposable elements could therefore provide a mechanism for transposition of the *orf1* gene to a new genetic locus. The genes for these transposable elements were disrupted and presumed to be non-functional (Dunn, 1996), perhaps indicating that this transposition event occurred at an ancient point in evolutionary history. This would fit with the hypothesis above that several further steps were then necessary to form the 130-kDa protein genes. Similar transposable elements have been found on either side of the Cry4A gene in Bti (Delécluse *et al.*, 1989), these elements are not disrupted and may indicate that they were involved in more recent transposition events. The existence of these transposable elements could explain how such similar genes as those described above are present in such a variety of host strains.

According to this hypothesis, the two gene arrangement must therefore have been the progenitor of the single-gene 130-kDa protein not *vice versa*. What it does not explain is why the short Cry11A protein should be efficiently expressed in a stable form without the aid of a chaperone, while the short Group 3 proteins are not. This would seem to be a retrogressive evolutionary step. It could however have occurred by one of two possible mechanisms. (i) A mutation or mutations in a *cry11A* progenitor caused the protein product to be especially susceptible to proteolysis leading to the formation of a cryptic protein gene whose product was not present in the inclusion. Insertion of the ORF2 protein gene immediately downstream of the damaged *cry* gene then allowed the synthesis of ORF1/ORF2 inclusions. Or (ii) insertion of the ORF2 protein gene downstream of a functional *cry* gene compensated for subsequent mutations in the ORF1 gene that increased its protease sensitivity. These mutations were not therefore deselected, eventually leading to an ORF1 product that was dependent on the ORF2 product for its stable expression.

Evidence for the first mechanism may come from the discovery that cloning the *orf1* and *orf2* genes of the *cry2A* operon with the cryptic *cry2B* protein gene allows synthesis of Cry2B inclusions (Crickmore *et al.*, 1994). *cry2B* may therefore be an example of the intermediary non-functional ORF1 protein gene from this hypothesis. This is however, a hypothesis based on a very limited data set. It is especially difficult to determine the

phylogenetic relationship between Cry11A and Cry20A. As more *cry* genes are identified and the groups are enlarged a clearer picture should emerge.

In order to be consistent with the hypothesis proposed above, one would expect *am1* and *am2* to be associated with transposable elements. The search for such elements in the DNA surrounding *am1* and *am2* is discussed in the next Chapter of this Thesis.

Eight

Chapter 8

Sequence Analysis of DNA Surrounding the am1 Gene

8.1 Introduction

Cloning *am1* and the second open reading frame *am2* involved the coincidental cloning of large regions of upstream and downstream DNA. Sequence analysis of these regions was carried out to determine the genetic context of the Bt ssp. *fukuokaensis am1* gene. This was thought to be important since, as is described below, there seemed a high probability of discovering additional open reading frames encoding proteins of interest.

A variety of Bt toxin genes have been found in association with transposable elements (Kronstad & Whiteley, 1984; Lereclus *et al.*, 1984). The presence of toxin genes on large (>30 MDa) self-transmissible plasmids and their close association with these genetically mobile elements has been thought to explain the widespread occurrence and distribution of δ-endotoxins; the suggestion being that conjugation may play a role in dissemination between naturally occurring *Bacillus* populations (Carlton & Gonzalez, 1985; Gonzalez *et al.*, 1981). Both Class I and Class II elements have been identified in Bt (reviewed by Mahillon *et al.*, 1994) and to date five insertion sequence (IS) elements; IS*231* (Mahillon *et al.*, 1985), IS*232* (Menou *et al.*, 1990), IS*240* (Delécluse *et al.*, 1989) ISBT*1* and ISBT2 and two transposons; Tn*4430* (Lereclus *et al.*, 1986) and Tn*5401* (Baum, 1994) have been characterised. In several cases two copies of an insertion element have been found to flank a *cry* δ-endotoxin gene forming a composite transposon (Menou *et al.*, 1990; Delécluse *et al.*, 1989; Bourgouin *et al.*, 1988; Kronstad & Whiteley, 1984). Such associations have been thought to enhance the spread of toxin genes within Bt subspecies leading to the natural production of new strains displaying novel δ-endotoxin combinations (Mahillon *et al.*, 1994). A Bt *cryII* gene has recently been identified in *B. popillae* (Zhang *et al.*, 1996) and may be an example of the transfer of a gene from one *Bacillus* species to another. Similarly, the identification of a transposable element upstream of a novel *cry* gene from *Clostridium bifermentans* ssp. *malaysia* (Barloy *et al.*, 1996) may indicate the first transposition-mediated transfer of a Bt gene to another bacterial species. It appears therefore that many δ-endotoxin genes are organised in transposon-like structures in which flanking insertion sequences might be the active elements in transposition. The structural characteristic suggests a possible mechanism for creating the diversity of δ-endotoxin genes found in Bt species.

The likelihood and significance of discovering transposable elements in the DNA surrounding *am1* and *am2* has been discussed in Chapter 7. In short, the similarity of *am1* and *am2* to toxin genes from other Bt subspecies (themselves associated with mobile genetic elements) indicates that transfer between these subspecies has taken place at some point in evolutionary history. In particular, the *cry4A* gene of Bt ssp. *israelensis* is flanked by two copies of the transposase IS*240* (differing by only 6 bp) (Bourgouin *et al.*, 1988; Delécluse *et al.*, 1989). *am1* and *am2* bear considerable similarity to *cry4A*, therefore it does not seem unreasonable that transfer of a *cry4A* type gene may have occurred, either from 17A to Bti or from Bti to 17A, at some point. If this had occurred then evidence for the transfer may still exist in the form of similar transposable elements.

Several open reading frames with significant similarity to previously identified transposable elements have recently been identified by Dunn (1996) during sequence analysis of the DNA surrounding the cloned *orf1* gene in Bt ssp. *fukuokaensis* 84-I. An IS*240* like insertion sequence was found to be located immediately upstream of *orf1*, whilst further upstream, an IS*150* like pair of open reading frames and the gene for a site-specific recombinase shown to be involved in Class II transposition formed what was possibly part of a novel Class II transposon.

In addition to transposable elements Bt toxin genes are often found in regions of DNA containing other toxin genes and/or accessory proteins (chaperones) sometimes as part of an operon. For instance *cry11A* from Bt ssp. *israelensis* is located in an operon along with the *cyt1A* toxin gene and two accessory proteins (*p20* and *p19*). Expression of the 20-kDa protein with *cry11A* in *E. coli* has been found to significantly improve net synthesis of Cry11A and promote crystal formation (Wu & Federici, 1995); this has also been shown to be the case for Cyt1A in Bt. As 17A has been shown to contain several crystal proteins (see Chapter 3), it is possible that one or more of these may also be located in an operon along with *am1*.

A final reason for believing that analysis of the surrounding DNA sequence might prove of worth was also provided by work carried out by Dunn (1996). Sequence analysis of DNA downstream of *orf1* identified a novel genetic locus Btvir. Characterisation of this region identified two genes *btcap1* and *btpk1* whose putative protein products were found to have homologues in other pathogenic micro-organisms. Btcap1 was found to display similarity to a protein essential for polysaccharide capsule biosynthesis in *S. aureaus* and also to an *E. coli* protein associated with the polysaccharide chain-length determinant of lipoploysaccharide (LPS). From the deduced putative promoter sequence it was predicted to be produced by vegetative Bt. Btpk1 was found to display similarity to a variety of eukaryotic-type Ser/Thr protein kinases. The identification of these two genes has important implications for the pathogenic mechanisms of Bt. The similarity of Bt to the more widely pathogenic *B. cereus* has promoted an interest in secondary factors produced by

Bt which may enhance the effect of the δ-endotoxin crystal and thus the insecticidal potency of Bt.

The toxicity of Bt ssp. *fukuokaensis* 17A crystalline inclusions to *Dacus oleae* was found to be enhanced on addition of spores (Yu *et al.*, 1991). Similar results have been shown with a variety of other insect larvae treated with a mixture of spores and crystals (Li *et al.*, 1987; Van-Nguyen, 1995). The implication is that if vegetatively growing Bt can invade the insect then other factors produced by vegetatively growing Bt may aid the organism in its observed virulence. The initial effect of the δ-endotoxin ingested by a susceptible larval insect together with a Bt spore, is to cause lesions in the host gut epithelium. Such lesions have been found to be either lethal or to result in weakening of the larvae. The subsequent release of nutrients from damaged host cells results in a lowered midgut pH, thus providing a suitable environment for spore germination (Dadd, 1975; Van-Nguyen, 1995). Germinating spores may then be able to invade the haemocoel. Evidence for this comes from the identification of spores in the insect haemolymph (Miyansono *et al.*, 1994). It is in this environment that expression of additional virulence factors would be effective. For example the immune inhibitor A produced by vegetative Bt specifically degrades the antibacterial proteins (attacins and cecropins) produced by the insect. Other non entomophathogenic *Bacilli* have not been found to express this protein (Lövgren *et al.*, 1990).

If *btcap1* and *btpk1* are indeed virulence factors, expression in vegetative cells may play a role in the overall pathogenesis of Bt. The Btcap1 protein may be involved in the formation of a capsule that surrounds the bacterial cell assisting evasion of host antibacterial mechanisms. Such a capsule may allow Bt to proliferate in the susceptible insect haemolymph and express the other secreted factors necessary for virulence. Btpk1 is predicted to play a more direct role in virulence in that if secreted, it could interfere with the cell-signalling apparatus of the insect larvae and thus contribute to the pathogenic mechanisms. The presence of *btcap1* and *btpk1* in 17A and other Bt strains has been indicated by PCR and Southern blot analysis using oligonucleotides designed towards these two genes (Kinchen, personal communication). With the evidence that vegetative growth is necessary for the full virulence of 17A (Yu *et al.*, 1991), the implication is that these and possibly other virulence factors may also be found in this related strain. If, as has been found with 84-I these factors form part of a 'virulence locus' then it is not unreasonable to expect that they may be found associated a crystal toxin gene, the major virulence determinant of the strain.

8.2 Sequencing Strategy for DNA Surrounding *am1*

Restriction analysis of the two 17A clones (pAM15 and pAM10) identified suitable restriction sites that could be used to generate subclones to facilitate the sequencing of DNA upstream and downstream of *am1*. The subcloning method is detailed in Chapters 4 and 5 and so is only outlined here. *Pst*I digestion of pAM15 generated four restriction fragments (see Figure 8.1), one of which (pAMsc8) contained *am1* and had already been subcloned and sequenced (see Chapter 5). The two smaller fragments were cloned into the *Pst*I site of the pGEM-3Zf(+) polylinker generating subclones pAMsc6 and pAMsc7 (Figure 8.1). The largest fragment already contained linearised vector from the original clone and was simply recircularised to form the fourth pAM15 subclone (pAMsc5) containing approximately 1.5 kb of 17A DNA (Figure 8.1).

As is described in Chapter 5, *Xba*I digestion of pAM10 generated four fragments. The largest fragment hybridised to DIG-labelled 17A.1 and contained linearised vector DNA, this was recircularised to form the subclone pAMsc1 containing approximately 4 kb of 17A DNA, most of which was duplicated in the subclone pAMsc8. The other three fragments were cloned into the *Xba*I site of the pGEM-3Zf(+) polylinker.

In total, this generated a library of eight subclones containing the whole of the DNA from the two original clones. Subclones pAMsc1 and pAMsc2 containing the genes *am1* and *am2* had already been sequenced and as this region overlapped with that of pAMsc8 there was no need to resequence this clone. In order to sequence the remaining downstream DNA, sets of unidirectional deletions were made in pAMsc3 and pAMsc4 using the modified nested deletions protocol described in Chapter 5. Deletions were made in both subclones using *Kpn*I to generate the nuclease resistant end and *Bam*HI to generate the nuclease sensitive end. To sequence the complementary strand of pAMsc4, further deletions were made cutting at restriction sites *Hin*dIII (nuclease resistant) and *Sal*I (nuclease sensitive). To sequence across the breakpoints between subclones, oligonucleotides were designed to the sequence approaching both ends of each subclone. These oligonucleotides were then used as primers in sequencing reactions on the clone pAM10. Similarly, oligonucleotides were used to obtain the complementary strand of any open reading frames of interest within pAMsc3. Resulting sequences were assembled using the bestfit program in the GCG package (Section 2.1.6) to generate approximately 5 kb of continuous sequence downstream of *am2*.

Sequencing the DNA upstream of *am1* present in clone pAM15 proved more difficult due to a lack of suitable nuclease resistant sites in the pGEM polylinker from which to generate the nested deletions. The only available nuclease resistant site not also present in the three subclones was *Sac*I and all attempts to generate completely linearised DNA using this enzyme were unsuccessful. In test digests using <0.1 µg/µl of DNA *Sac*I cut to completion and a single band was seen when the DNA was subject to agarose gel

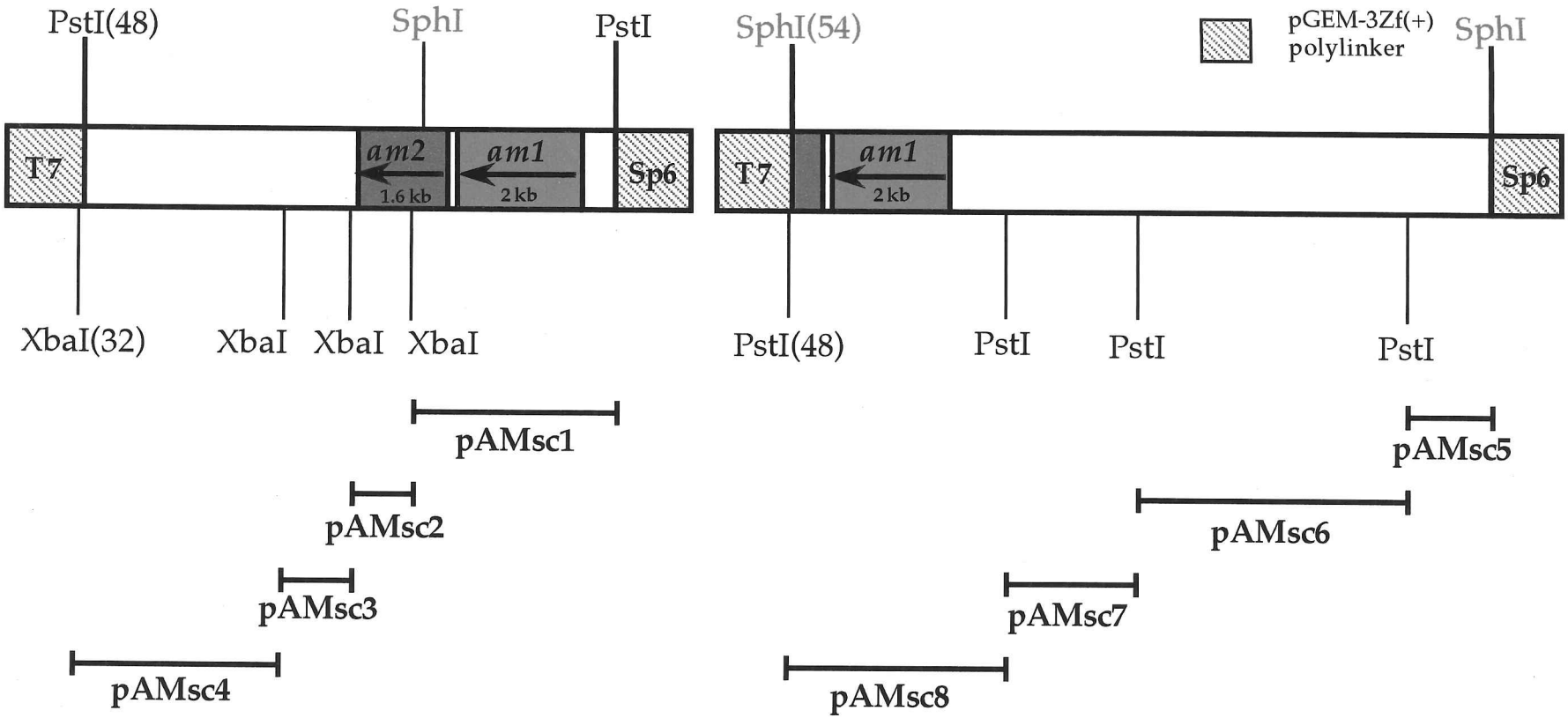


Figure 8.1 Construction of subclones from pAM10 and pAM15.
The bars pAMsc1 to pAMsc8 represent the fragments of DNA cloned into the vector pGEM-3Zf(+) in the construction of the subclones pAMsc1 to pAMsc8. The subclones contain inserts of 17A DNA of the following sizes:-

pAM10 subclones	size of insert (bp)	pAM15 subclones	size of insert (bp)
pAMsc1	3779	pAMsc5	1496
pAMsc2	969	pAMsc6	5274
pAMsc3	1218	pAMsc7	2153
pAMsc4	3592	pAMsc8	3621

electrophoresis. But, when the DNA concentration was raised to 1 µg/µl for use in the deletions experiment, multiple bands were present on a test gel indicating incomplete digestion of the plasmid DNA. *Sac*I is particularly salt sensitive so attempts were made to reduce the salt present in the reaction mixture first by reducing the concentration of One-Phor-All buffer to 0.5x and then by using a low salt buffer (Buffer L (10x): 100 mM Tris pH 7.5, 100 mM $MgCl_2$, 10 mM DTT). Neither of these conditions were successful in generating the completely linearised plasmid DNA needed as a substrate for Exonuclease III in the deletions reaction.

Deletions of the three *Pst*I subclones were attempted again following Procedure B from the Nested Deletions kit protocol. This involves linearisation of the DNA using a restriction enzyme to generates a 5' overhang that is then rendered resistant to exonuclease III digestion by "filling-in" with thionucleotides using Klenow fragment. *Bam*HI (pAMsc5 and pAMsc6) or *Eco*RI (pAMsc7) enzymes were used in the initial linearisation. After this had been end-filled, *Sal*I (pAMsc5 and pAMsc6) and *Xba*I (pAMsc7) were used to generate the nuclease sensitive sites. Nested deletions were then carried out using the Exonuclease III enzyme but again these proved unsuccessful. When DNA from the reactions was run on a gel, two bands were visible at each time point for pAMsc6, one corresponding to the deletion but the other to full length linearised DNA. No deletions were evident at all for the two smaller subclones. This could have been due to incomplete cutting by the second enzyme but both of these had been used successfully in previous deletion experiments with the pAM10 subclones. The proximity of the *Bam*HI and *Xba*I sites on the pGEM polylinker may have meant that the second enzyme was unable to cut to completion but this should not have been the case for the *Eco*RI and *Xba*I sites. A more likely explanation was that, despite heat treatment and DNA precipitation, carry over of Klenow and dNTPs into the second digestion reaction had caused filling in of the second 5' overhang.

To try and prevent possible carry over of Klenow fragment the volume of Klenow fragment used in a subsequent experiment was reduced by half and the DNA phenol extracted before use in the second digestion. In this experiment deletions were obtained for pAMsc5 but again full length bands were visible in all timed samples for pAMsc6 and no deletions were evident for pAMsc7. Despite repeated attempts, deletions were never obtained for subclones pAMsc6 and 7 and these were eventually sequenced by "gene walking" as previously described.

Sequencing across the breakpoints between the pAM15 subclones was again carried out using oligonucleotides and the resulting sequences compiled as before to generate 9.8 kb of continuous upstream sequence.

8.3 Sequence Analysis of DNA Downstream of *am1*

Analysis of the nucleotide sequence revealed the presence of several open reading frames greater that 50 codons. These included four major open reading frames greater than 150 codons. Figure 8.2 shows the complete 4900 bp of DNA sequence from clone pAM10 downstream of *am2*. EMBL and SWISSPROT data bases were searched for sequences similar to both the DNA and protein sequence of any predicted open reading frame greater than 50 codons. Only the first major open reading frame (*T1*), thought to run from nucleotide 1240 to 549 and orientated in the opposite direction to *am1*, was found to show any significant similarity to previously identified proteins. *T1* was found to show sequence similarity to a variety of bacterial transposases; most notably over 70% amino acid sequence similarity to an IS*214*-II transposase gene previously identified in *Lactococcus lactis* ssp. *lactis* K214 and *Enterococcus faecium* HM1073. *T1* is discussed in greater detail in Section 8.5.

The second major open reading frame in the downstream section of DNA encoded a putative 227 amino acid protein (OrfB) again orientated in the opposite direction to *am1*. This *orf* was thought to run from nucleotide 2343 to 1660. Database searches carried out on OrfB did not reveal significant similarity to any known proteins. A database protein motif search called MotifFinder (see Section 2.1.6) was used to search a library of sequence motifs for patterns in the OrfB amino acids sequence. This search facility is designed to help understand the biological meaning of query sequences. If any patterns are found MotifFinder then searches both protein sequence (Swiss-Prot or PIR1) and structural databases for those sequences and structures that contain the patterns. The results are given in graphical representations where the location of the pattern on the 3-D structure can be viewed in relation to known biological function sites annotated in the sequence data bases. Both Motif dictionary and prosite data bases were searched using this facility but no motif was found to be matched to the query sequence.

1065 base pairs downstream of OrfB lay a third major open reading frame OrfC. OrfC is 260 amino acids long although there is no ATG start codon present until 156 amino acids into this open reading frame. Overlying the end of OrfC and reading in the opposite orientation is a 99 amino acid open reading frame with the first ATG start codon 16 amino acids into the *orf*. BLAST database searches of both of these amino acid sequences revealed no significant similarity to any known proteins. Nor was any structural Motif identified by MotifFinder.

At the end of the cloned DNA was the partial sequence of a final open reading frame (OrfD) with a putative start site past the end of the clone. A BLAST database search revealed some similarity to integrase/recombinase proteins from a range of different bacteria. Most notably XERC and XERD from *Haemophilus influenzae* and *E. coli*. XERC and

```
  G   C   T   C   N   Q   G   H   N   S   G   C   T   C   N   Q   G   T   N   R
GGT TGC ACA TGT AAC CAA GGG CAT AAC TCT GGC TGT ACA TGT AAT CAA GGA TAT AAC CGT 60

  *
TAA CGA TTC TAA ATA AGA ATC AAC ATC ATT GCG AAA TAT AAA AAC CTA CTC ACA AAA TCT 120

ATT GCA TAT CAT AAC ATA AGC TTT ACA AAT AAC TGA CAT ATT CTA GAA GAG GTC TCC TTA 180
                                    L   Y   S   V   Y   E   L   L   P   R   R   L

ATT TTA AAA TAA GGA GAT CCT TTT CGT TTC CCC AAT ATT GAT TAA TGA AAA TAA GAT ACA 240
 K   L   I   L   L   D   K   E   N   G   W   Y   Q   N   I   F   I   L   Y   L

ATT TAT GAG CAT TTA TTA GAT TCA TTT TTT ACA ATT TCT TTG AAT AGC AAA TAC AAG ATC 300
 K   H   A   N   I   L   N   M

AATCTAACCA ATTAAAGAAA AGCTTTCTTC GTTTGTTTTA TTTACTTACT AAAAATGGAC 360
GAAAATCACA TCATTATCAT AAGAAAATTA TATTTGATAT CTCAAGTATT GCATACGAAC 400
AACATCTATC AAAATAAAAA ATGAAATATT TTTAAATTTT TAACCCGTAA TGGCTATTTA 460
ATATGCCATT CTTTCCGTAC CCCATAACTA CTGGAGGTTC TGTTGCAAAG TTTCTTTACA 520
                                    ---------------->

GATAAAAAGT TGGTATGATT GCAGATAG TGC GAT CGT TAG AAA TTG TTG TAA TTC 580
                              ACG CTA GCA ATC TTT AAC AAC ATT AAG
                              A   I   T   L   F   Q   Q   L   E

ATT GTA CGT CGA AAA AGC GAA GTT TGG TTG TTG CAA ACT TCG TTT TCG TTT ATA TAT AGC
TAA CAT GCA GCT TTT TCG CTT CAA ACC AAC AAC GTT TGA AGC AAA AGC AAA TAT ATA TCG
N   Y   T   S   F   A   F   N   P   Q   Q   L   S   R   K   R   K   Y   I   A

ATG AAT TGT TTC GAT TCC TTT CAA GGT GCG TGA AGC ATG GCG AAG ACT TTG AAA TCC TGC 700
TAC TTA ACA AAG CTA AGG AAA GTT CCA CGC ACT TCG TAC CGC TTC TGA AAC TTT AGG ACG
H   I   T   E   I   G   K   L   T   R   S   A   H   R   L   S   Q   F   G   A

GGA TCT GGA AAA ACG TCG CTT GAT ATG TCT ATG ATC CTG TTC AAT GAG ATT ATT CAA ATG 760
CCT AGA CCT TTT TGC AGC GAA CTA TAC AGA TAC TAG GAC AAG TTA CTC TAA TAA CTT TAC
S   R   S   F   R   R   K   I   H   R   H   D   Q   E   I   L   N   N   L   H

CTT GAT TGT ACA ATG AGT GGT ATG CTT ATA AAA GCC TTG TTC TCT CAA TTT ATT GAA CGC 820
GAA CTA ACA TGT TAC TCA CCA TAC GAA TAT TTT CGG AAC AAG AGA GTT AAA TAA CTT GCG
K   I   T   C   H   T   T   H   K   Y   F   G   Q   E   R   L   K   N   F   A

ACA AAG TAG CGC GGG CGC TTT GTC TGT TGT GAG ATC CGT TGG TTC TCC AAA TGT TTT CAC 880
TGT TTC ATC GCG CCC GCG AAA CAG ACA ACA CTC TAG GCA ACC AAG AGG TTT ACA AAA GTG
C   L   L   A   P   A   K   D   T   T   L   D   T   P   E   G   F   T   K   V

TAG TCT TTT CAT AAA GGC ATA TGC AGC TTG ATG ATC TCG TTT TTT GCG AAG TTG AAT ATC 940
ATC AGA AAA GTA TTT CCG TAT ACG TCG AAC TAC TAG AGC AAA AAA CGC TTC AAC TTA TAG
L   R   K   M   F   A   Y   A   A   Q   H   D   R   K   K   R   L   Q   I   D

CAA TGT ATG TCC ATC TTT ATC GAT TGC TCG ATA GAG ATA ACA CCA TTC GCC TTT CAC TTT 1000
GTT ACA TAC AGG TAG AAA TAG CTA ACG AGC TAT CTC TAT TGT GGT AAG CGG AAA GTG AAA
L   T   H   G   D   K   D   I   A   R   Y   L   Y   C   W   E   G   K   V   K
```

Figure 8.2 Nucleotide sequence of the 4900 bp of Bt ssp. *fukuokaensis* DNA downstream of *am2*.

The end of *am2* is shown from base 1 to 60. The final TAA stop codon is indicated by an asterisk. The sequence of the putative Insertion Sequence *T1* is shown from base 1240 to 549. The position of a 16 base pair inverted repeat sequence is indicated by opposing arrows.

```
GAT ATA GGT TTC ATC CAG GTG CCA AGA TAG CTG TGC GGA TTT GTT TTT CTT TTT CCA AAT 1060
CTA TAT CCA AAG TAG GTC CAC GGT TCT ATC GAC ACG CCT AAA CAA AAA GAA AAA GGT TTA
I   Y   T   E   D   L   H   W   S   L   Q   A   S   K   N   K   K   K   W

TTG ATA CAT CTG ATT TCC ATA TTC ATG TAC CCA ACG CAT AAT CGT TGT CGG ATG AAC AGA 1120
AAC TAT GTA GAC TAA AGG TAT AAG TAC ATG GGT TCG GTA TTA GCA ACA GCC TAC TTC TCT
Q   Y   M   Q   N   G   Y   E   H   V   W   R   M   I   T   T   P   H   V   S

GAT GCC ACG TTC CCT CAA AAT TTC AGA TAC ATC ACG ATA ACT TAA AGA AAA ACG ACA GTA 1180
CTA CGG TGC AAG GGA GTT TTA AAG TCT ATG TAG TGC TAT TGA ATT TCT TTT TGC TGT CAT
I   G   R   E   R   L   I   E   S   V   D   R   Y   S   L   S   F   R   C   Y

ATA GCC AAC GGC TAC CAA AAT AAT ATC TTT CTT GAA TTG TTT TCC TTT AAA ATA TCT CAT 1240
TAT CGG TTG CCG ATG GTT TTA TTA TAG AAA GAA CTT AAC AAA AGG AAA TTT TAT AGA GTA
Y   G   V   A   V   L   I   I   D   K   K   F   Q   K   G   K   F   Y   R   M

GTA TCA TGC CCC TCG AGT CAT TTT CTT TAA ATT TTA ACT TTG CTT AGA CAT CTT TGC AAC 1300
                                                          <----------

AGA ACC TAA TTT TTT TCT TAT TGA ACT AAC GGG GCA GTT TAG TTT AAT AAC AAA AAA ATC 1360
---------                                               *   Y   C   F   F   R

TAG AAC TAA TGT ATA ATC CAT TAT GAA TAT TAA AAA AGT GGA GGG TTG CAT ACA CAA CAC 1420
 S   S   I   Y   L   G   N   H   I   N   F   F   H   L   T   A   Y   V   V   G

CTC CTA TTC TTT TTC TGC CGC ATT TAT GGG CTT CTT GTG TAG CTT TGA AAA TAT AGC GCG 1480
 G   I   R   K   R   G   C   K   H   A   E   Q   T   A   K   F   I   Y   R   S

AAA TCA TGA AAT GTG TTC TTA TAA AGG GAA AAA CTT CCC ATA GCT CAT CCC GAT TGC TCT 1540
 I   M

GAT GCA AAA GCG TAG AAT TTT TAT ATT AGC TGA ACT TTC ACC CTA ATG TCT CCT GTA ATA 1600
AAA AAA GCC GCT TAA TAA GCG ACT CAT ACA TAG GTC AAC GCC GAT AAT TCG GCG GCT GCT 1660

TAA CGA GCT CTA TTT AGT TTC TCC TTG AAT TCC TTT AAT GGT CGC TTA CGA CCA TCA TGA 1720
  R   A   R   N   L   K   E   K   F   E   K   L   P   R   K   R   G   D   H   Y

TAA GGT TTA TCA TAA ACA ACA TCT TCT GAT ATC ATG ATT TGC CTC ATT AAA GCA TGA TAG 1780
  P   K   D   Y   V   V   D   E   S   I   M   I   Q   R   M   L   A   H   Y   V

ACA GCA GGA TGA AAA ACA ATG GAG TCT TGG GGG TAT CGT GCA AAT ACC CAT TCA TCA TAA 1840
  A   P   H   F   V   I   S   D   Q   P   Y   R   A   F   V   W   E   D   Y   D

TCG ATA TCT TTG CAA GGG TCG TAT TTA TCA ACT AAC TCA AAG GTA TCG AAT TTT TTT CCG 1900
  I   D   K   C   P   D   Y   K   D   V   L   E   F   T   D   F   K   K   G   N

TTA CCA ATC TTT TCA ATC ATT TCT TCC ATT ATC TCC ACA TGA GTA TCA GGT CCG TAG GGG 1960

TCT TTA AGA AAA GCA TCA AAG AGT TCG TCG AAT GTT GGA GCC CAG ATA TGT AAC TTT TTT 2020
  K   L   F   A   D   F   L   E   D   F   T   P   A   W   I   H   L   K   K   E

TCT TTA GGA AAA TGA CTT TCC ATT ACA AAG ATC AAT GAG ACA CAT CGC CCT TCC CAG CCG 2080
  K   P   F   H   S   E   M   V   F   I   L   S   V   C   R   G   E   W   G   C

CAT CTC GTG ATT AAT ACT TTA TTG TCA TAG GTG TCG CCT CCA ATA ACA AGA ACT TCT TCG 2140
  R   T   I   L   V   K   N   D   Y   T   D   G   G   I   V   L   V   E   E   I
```

Figure 8.2 continued.
Nucleotide sequence of the 4900 bp of Bt ssp. *fukuokaensis* DNA downstream of *am2*. The sequence of the putative Insertion Sequence *T1* is shown from base 1240 to 549. The position of a 16 base pair inverted repeat sequence is indicated by opposing arrows. The sequence of *orfB* is shown from base 2343 to 1660.

```
ATA TTT TCG TTT AAT GTC ACG TTA CCA TGC TCA TCA ATG TCC TGT TCT CGT TTA TCA TCG 2200
  N   E   N   L   T   V   N   G   H   E   D   I   D   Q   E   R   K   D   D   V

ACC ACT ACT TGA CCT GTA TAA TCG AAA CGG GGT TTA TCC AGA TAT TGA AAT TTC ATA TCC 2260
  V   V   Q   G   T   Y   D   F   R   P   K   D   L   Y   Q   F   K   M   D   W

CAC GGG TCC AAC GGT ACT CTA TCT ATA ACA CTC TCA CAT ATG AAC CGA GGA AAT CCC TTC 2320
  P   D   L   P   V   R   D   I   V   S   E   C   I   F   R   P   F   G   K   E

TCT GAT AAA TCA TCA TAA TTC ATT TTT TTT ACC CCT CCG ATA AAA GTA TTT AAG GAA CGT 2380
  S   L   D   D   Y   N   M

CTACATTAAT ATTAAACTAT GGAAATGATT GTCACAATTT AATTTTACAG AGTATCGTTA 2440
CAATGGCAGA TGAATGAATT TACATCTAGG TTTATCACTC CTAAAGGGAG GAGAATCAAA 2500
TGCAGGTTGT CAAAGAAGAG CTATAAAAGT CTTAAAAGAA CTATTTTAGA TACTTTTGGA 2560

                                                M   M   G   Y   F
GATAAAAAAG TAGTGATAGT TACTCACGGG GCTGTTATGA CCTAA ATG ATG GGT TAT TTC 2620

D   S   A   Y   D   L   N   F   L   H   S   T   S   K   P   D   I   Y   R   I
GAT AGC GCT TAT GAC TTG AAT TTT TTA CAT AGT ACA TCA AAA CCT GAT ATT TAC AGA ATT 2680

E   F   K   E   M   E   L   V   N   V   Q   R   I   W   G   V   N   C   *
GAA TTC AAA GAA ATG GAA TTA GTT AAT GTT CAA AGA ATA TGG GGT GTG AAT TGT TAA CCA 2740

ATAATAATTA TTTTTTGTTA TTGAACTAAA GGGTCCGTTA GTTTAAGAAG AAAGGACATA 2800
ATGTAGCTTT GAAATAAAGT CGTACTGTTT ATAAAAAAGA TGTACCGTTT ATAAAAAAGT 2860
TTCACCGCTT TATCCCTTCA ATTAAAGAAT GCTAAGAGAG CTTTGATCAA ACTTTTTTAT 2920
AAAAATCTAA TACGTATTTA ATAACATTAA CTCGATTTTG AACAATTTTT TTCAAAATGT 2980
GAGGTTTTCT TCTATTAGTA AATCAACATT TGCAGCCTAC TTTTGATCAA GTTTTATTCC 3040
AAACCTACCA TAGGAATTAC AAATACTTTT TGTGAAAAAA TATTATATCC ATCTTTATAA 3100
AAAGTATAAT AATTAAATAA TGAAAGCCGA TTTAATATCT AAAAAGGGAG AACCTAATGT 3160
CCTATTTTAT AGTTTTGGAT ACCAATATAA TATCTAATGA TTTCTTTTTT AAATCTGCCG 3220
ATATGAAAAA ATTACCAAAT CCCACAACAT AAACCAGTAA AACTTTGTTT CACAAAGTTT 3280
AACTATCATG AAATCCTAAA AAATTTAAGG TGAAATTAGA CTTCTGATTA AAACAGTAAA 3340
GGTACTAGAG CTGATTTAAT AAGGTTGGAA GCATCCGAGA TAATTAATTT TGATAATTTA 3400

          Y   I   C   E   K   Y   K   N   T   F   D   G   I   I   G   E   N   N
AAA GCT GAT ATA TTT GTG AAA AAT ATA AGA ATA CTT TTG ATG GAA TAA TTG GAG AAA ATA 3460

  I   K   I   I   D   F   P   K   S   E   D   V   T   E   T   I   S   L   K   Y
ATA TTA AAA TCA TTG ATT TTC CTA AAT CGG AGG ATG TTA CGG AGA CAA TTT CAC TTA AAT 3520

  F   N   N   E   K   P   F   D   E   N   K   V   S   F   Q   D   A   I   I   W
ATT TTA ATA ATG AGA AAC CGT TTG ATG AAA ATA AAG TTT CTT TTC AAG ATG CAA TTA TAT 3580

  E   S   I   V   E   Y   C   R   D   N   E   P   D   Q   V   A   F   I   S   N
GGG AAA GTA TAG TGG AAT ATT GTA GAG ATA ATG AAC CAG ATC AAG TAG CAT TTA TCT CAA 3640

  N   H   K   D   F   A   N   K   D   K   N   D   I   H   E   D   L   A   G   D
ATA ATC ACA AAG ATT TTG CTA ATA AAG ATA AAA ACG ATA TTC ATG AGG ATT TAG CAG GGG 3700

  I   Q   N   L   S   Y   Y   N   S   L   S   A   F   L   E   N   E   E   E   N
ATA TTC AAA ATT TAT CGT ATT ATA ATT CAT TAT CAG CAT TTT TAG AAA ATG AAG AAG AAA 3760

  L   R   D   Y   F   N   D   N   F   E   Y   D   E   E   S   L   I   N   D   L
ATT TAA GAG ATT ACT TTA ATG ATA ACT TTG AGT ACG ATG AAG AAT CAT TAA TAA ATG ATT 3820
```

Figure 8.2 continued.
Nucleotide sequence of the 4900 bp of Bt ssp. *fukuokaensis* DNA downstream of *am2*. The sequence of *orfB* is shown from base 2343 to 1660 and that of *orfC* from 3409 to 4183.

```
  K   S   F   C   E   R   N   N   F   I   Q   D   T   I   D   D   M   L   F   N
TAA AAT CAT TCT GTG AGA GAA ATA ATT TTA TTC AGG ATA CTA TTG ATG ATA TGC TAT TTA 3880

  D   H   F   E   G   E   Y   F   S   G   W   G   T   D   G   Y   I   E   E   Y
ATG ATC ATT TTG AAG GTG AAT ATT TTA GTG GAT GGG GTA CTG ATG GGT ATA TTG AGG AAT 3940

  D   I   T   I   S   E   V   S   L   D   I   E   D   N   A   M   L   I   S   F
ACG ATA TTA CTA TAA GTG AAG TTT CTT TGG ATA TTG AAG ATA ATG CTA TGT TAA TAT CTT 4000

  D   I   Q   A   D   V   S   F   S   I   E   T   K   D   P   T   Y   E   R   G
TTG ACA TTC AAG CCG ATG TGA GTT TTA GTA TTG AAA CAA AAG ATC CGA CAT ATG AAC GCG 4060
                                    Y   Q   F   L   L   D   S   M   H   V   R

  D   P   G   D   G   L   I   S   E   S   S   S   T   Q   I   L   I   H   S   N
GAG ATC CTG GCG ATG GAT TGA TAT CAG AAA GTA GTA GTA CAC AAA TAT TGA TAC ATA GTA 4120
L   D   Q   R   H   I   S   I   L   F   Y   Y   Y   V   F   I   S   V   Y   Y

  V   T   Y   L   L   E   D   E   E   L   I   D   Y   V   E   L   D   R   E   F
ATG TAA CTT ACT TAT TAG AGG ATG AAG AAC TAA TAG ATT ACG TTG AAT TAG ATA GGG AAT 4180
H   L   K   S   I   L   P   H   L   V   L   L   N   R   Q   I   L   Y   P   I

  C   *
TTT GTT AGC ATG CTG ATT ATA TAA ATA TTT CCT TAA TGG GTA TAT TTT ATT TAG CGT GAA 4240
K   N   A   H   Q   N   Y   L   Y   K   R   L   P   Y   I   K   N   L   T   F

TTT GAA ATA GAG TGA CGA TAT TTG TAA GAA AGA CAT AAA GTT TAT AAA AAA GAT GTG ATG 4300
K   F   Y   L   S   S   I   Q   L   F   S   M

CTTTATCCCT TCAGGTAAAG GGCTTTAAGT GGGGTTTAAT CAAACTTTTT TATAAAAACT 4360
AACCTTAAAT ATGCTGGATA AGAATCCATT TTGATCAATC TTTTTCAAAA TGGATTCTTT 4420
TTATTTACCG TAGAATACGG GTTAACACGA TATTTTTAAT CAAACTTTAT TTCAAAGCTA 4480
CATTTATGTG GACTGCGTTT CTCATCGAAT GCACTTGAAT ATTTGTCACA TAACTTCTGT 4540
ATAGCTCTAA CAGTTAACTG TGTATATTTC CTACGAAGAG TAAGTCTTCA TCTTGAGGGC 4600

AGT TGT ATT TAC TAG GAC GCA ACC CTA ATG TAT TCC TGA ACA TCA TCT AAG CAA GAG CGA 4660
            G ATC CTG CGT TGG GAT TAC ATA AGG ACT TGT AGT AGA TTC GTT CTC GCT
            *   S   A   V   R   I   Y   E   Q   V   D   D   L   C   S   R   T

GTG GCT AAT ATG GAA TAT CTT TTA TTT CCT TTA CGT GTC ACT TTC ACG GTT CCC TGA CGG 4720
CAC CGA TTA TAC CTT ATA GAA AAT AAA GGA AAT GCA CAG TGA AAG TGC CAA GGG ACT GCC
  A   L   I   S   Y   R   K   N   G   K   R   T   V   K   V   T   G   Q   R   F

AAG TTT ATA CTG GAT ATA GTA AGG CTA GCA ACC TCA GAA ACT GGA AGA CCT GTT CCT AAA 4780
TTC AAA TAT GAC CTA TAT CAT TCC GAT CGT TGG AGT CTT TGA CCT TCT GGA CAA GGA TTT
  N   I   S   S   I   T   L   S   A   V   E   S   V   P   L   G   T   G   L   I

ATG AGG CTT ATA ATG GCT ATA TCT CTT TCT TTA TCG CGC TGG AAA AAA CTA AAT TTT CTT 4840
TAC TCC GAA TAT TAC CGA TAT AGA GAA AGA AAT AGC GCG ACC TTT TTT GAT TTA AAA GAA
  L   S   I   I   A   I   D   R   E   K   D   R   Q   F   F   S   F   K   R   P

GGT GAT ATA TCT TTC AAC ATT TCT TCG TAA TCA TTT CCT AAA AAT TGA AGA AAA GCT ACA 4900
CCA CTA TAT AGA AAG TTG TAA AGA AGC ATT AGT AAA GGA TTT TTA ACT TCT TTT CGA TGT
  S   I   D   K   L   M   E   E   Y   D   N   G   L   F   Q   L   F   A   V   D

TCA TCC TTT TCG TTA AAT ATC ATG TTT GCA ACG TCA TCT GCT CTT GCT GC 4950
AGT AGG AAA AGC AAT TTA TAG TAC AAA CGT TGC AGT AGA CGA GAA CGA CG
  D   K   E   N   F   I   M   N   A   V   D   D   A   R   A   A
```

Figure 8.2 continued.
Nucleotide sequence of the 4900 bp of Bt ssp. *fukuokaensis* DNA downstream of *am2*. The sequence of *orfC* is shown from 3409 to 4183. The partial sequence of a putative integrase/recombinase *orfD* is shown from base 4950 to 4613.

XERD participate jointly in the site-specific recombination of transposases by catalysing the cutting and rejoining of the recombining DNA molecules. The presence of OrfD is therefore likely to be linked to the presence of the putative insertion sequence T1. OrfD is discussed further in Section 8.5.

Several smaller open reading frames were also identified, the translations for which are shown alongside the sequence in Figure 8.2. Database searches with the amino acid sequence of these potential Orfs did not reveal any significant similarity to any known proteins, but, as the databases expand the identity of proteins encoded by some of these open reading frames may be revealed.

8.4 Sequence Analysis of DNA Upstream of *am1*

A six frame translation of the sequence upstream of *am1* (obtained from pAM15) was also examined for the presence of open reading frames of over 50 codons. The DNA and amino acid sequences of any such *orfs* were then subjected to BLAST data base searches for similarity to previously identified proteins. The amino acid sequence of these *orfs* is shown above the nucleotide sequence of the upstream region in Figure 8.3. Nucleotide number 1 is that which is furthest away from *am1*.

As was found with the downstream DNA, several major open reading frames of over 150 amino acids were identified with no obvious similarity to any known protein sequence (*orfF* [2311-2955], 164 amino acids; *orfG* [3578-4344], 256 amino acids; *orfI* [8328-8991], 221 amino acids). The first open reading frame to show any significant similarity was a small, 222 base pair open reading frame *orfE* shown between 1439 and 1659 base pairs in Figure 8.3. *orfE* encodes a putative 74 amino acid protein (OrfE) which shows significant homology to the amino acid sequence of α/β-type SASP (small acid-soluble spore protein) from a wide range of sporulating bacteria. SASP have an important role in the resistance of the spore to ultra-violet radiation and are degraded on germination to provide a source of free amino acids essential for spore outgrowth. SASP are ubiquitous *Bacillus* proteins, but the identification of a SASP gene on a plasmid is relatively unusual. The functional roles of SASPs and analysis of *orfE* are discussed in further detail in Section 8.6.

Just over 3 kb upstream from the start of *am1* and orientated in the opposite direction (7002 to 6585 in Figure 8.3) is an open reading frame (*am3*) encoding a 139 amino acid putative protein whose N-terminus shows almost complete homology to that of AM1. A Clustal PAM250 alignment was carried out to determine the degree of similarity between this protein and the start of the AM1 sequence (The first 150 amino acids of AM1 were used). The first 150 amino acids of Cry4A, Cry4B and Cry10A were also included in the alignment.

```
CCTTGAACAA GGAAATGCGG ATCTATATCA TTTAATGCAA ACACTAGGTC ACGAAAATAT 60
CCAAACAACA AAAATCTATT TAGAGAAGCA TATGAAAAGA AAGAATAATG TGGGGACTTC 120
ATTTGCGGAT ATGTTGATGT AAGATTTTAA TATAAGAACC TAATTTTATT ATGTGGTTTT 180
TATATAATTA ACAAGGGTTA TATTTGTATG AAAAAACAAT AAGAAAGTGA TTCCTATTGA 240
TAAAAAATAT ATTCAATCAA AAGTACTTTT GTAGTTAAGT TTAGTGACAG CTTGTAAATT 300
AGGAGCTACA CTTGGAATAG ATACATTAGC ACGTGGGATT GGCTCTGTCA GTGGAGAGAG 360
CCAATCCCTC AATGATTAAT AGTTGAAGCA AACGGATTTT TAGAATTAGG TATTTTCCTA 420
ATTAATTTCA TTGCTATTTC CTATTTTGGG GGTGGGAGTC TTACTGTTTA ATTTGAGGCT 480
TAGGGAAATA AAAGAGCGTG TTTAAAGCAG GAAAACAGGC TTTAAAAGCT TAAATAAGAA 540

          M   N   V   F   L   D   E   K   V   G   E   L   G   C   L   I   L   A
TTA AAA GTA TGA ATG TAT TTT TGG ATG AAA AGG TGG GGG AAT TGG GAT GTT TAA TTT TAG 600

  F   M   F   A   V   S   V   Y   F   L   V   F   D   N   P   Y   I   T   I   F
CAT TTA TGT TTG CGG TTA GCG TAT ATT TCC TAG TTT TTG ATA ATC CTT ATA TAA CTA TAT 660

  I   I   F   L   I   M   V   G   K   S   L   Y   D   N   R   K   N   R   E   *
TTA TTA TAT TTC TCA TCA TGG TAG GGA AGT CAC TAT ATG ACA ACA GAA AGA ACA GGG AAT 720

AAT CAG GGA TAC AGA GAG AGT ATG TGA GTA AAA TAG TGA AAT ATT GAT ACA TTT TAA TCA 780

                            M   I   F   E   N   V   S   T   M   G   A   L   V
TAA AGT AGG TGA TGT TCG AAA ATG ATT TTT GAA AAT GTT AGT ACG ATG GGT GCG CTT GTT 840

F   L   F   L   M   I   Y   L   A   T   D   P   K   D   V   S   L   L   T   I
TTT CTA TTT TTG ATG ATT TAT CTT GCG ACG GAT CCT AAA GAT GTA AGT TTG CTT ACA ATC 900

P   A   Y   F   G   G   M   W   I   T   N   W   L   N   E   N   G   F   Q   G
CCA GCT TAC TTC GGT GGA ATG TGG ATC ACT AAT TGG CTT AAT GAA AAT GGA TTT CAA GGT 960

T   F   M   Y   M   S   W   I   V   V   Y   V   I   L   M   F   Y   L   I   F
ACC TTT ATG TAT ATG AGT TGG ATA GTA GTC TAC GTC ATT TTA ATG TTT TAC CTT ATA TTC 1020

A   S   I   R   L   G   I   R   N   I   K   Y   I   K   E   K   I   R   K   R
GCT TCG ATT CGT TTA GGT ATA CGT AAC ATA AAA TAT ATT AAA GAG AAA ATA AGA AAA CGT 1080

R   A   I   K   K   *
AGA GCT ATA AAA AAA TAA CGT AAA GAG AGA CCG TCC TCA CGG GCG GTT TTA TTG TTA TAC 1140

TGAAGTGCAA TGTAGATTAT TTAACATTTG AGGATGGTTA ATTATAGAAT ACTCAAGAAT 1200
CTACACATTA AGAAACGGCC AAGAAAAATC AATAAAGGTA GAGTCATGGC GTTCATTTCG 1260
AGAAGAAATG AACGTATTAG GAATTCAGGA TTCAGATATT TTTCAAGTTC AGTTAGTTGA 1320
ACATGGGAAA GAGAGAACGA AAAAGAATTC TCCTAAAAAT CCTAGAAGAA AATAGTCTAC 1380
                                                                    M
TTA CGT ATA AAT ATC ATA ATT AAG CAC ATG TTA TAT ACA AAT ATA GTG GAG GTG CAA TTA 1440

  N   I   Q   R   N   E   S   K   S   S   T   N   E   I   F   I   S   A   A   A
TGA ATA TTC AAA GAA ACG AAA GTA AAA GTA GTA CGA ATG AAA TTT TTA TTT CTG CTG CAG 1500

  S   A   I   E   Q   M   K   Y   E   I   A   R   E   L   G   V   T   L   G   P
CAA GCG CTA TTG AGC AAA TGA AGT ATG AAA TTG CCC GTG AAT TAG GAG TTA CAC TTG GAC 1560

  D   T   S   S   R   A   N   G   S   V   G   G   E   I   T   K   R   L   V   R
CTG ATA CAT CAT CAC GTG CAA ATG GCT CTG TCG GTG GAG AAA TTA CAA AGC GCT TAG TTC 1620

  M   A   E   E   Q   L   T   G   Q   Y   R   L   H   *
GAA TGG CTG AAG AGC AAT TAA CAG GAC AAT ACA GAT TAC ACT AAG ATA AAA AAT AAA AGG 1680
                                                          =============

AAAAAGGTCC CCTAAGACGA ATCTTTTTCC TTTTATTTGC AAAAAATCCG AAAATCATTT 1740
=========>          <======================
```

Figure 8.3 Nucleotide sequence of the 9800 bp of Bt ssp. *fukuokaensis* DNA upstream of *am1*.

The sequence of the putative SASP gene *orfE* is shown from 1439 to 1659. A putative RBS is underlined and a region of dyad symmetry typical of SASP genes is marked by opposing arrows.

```
TTA AGA ACC AAA AAA CGA CTG CTC AGA TTC ATT TGT AAA AGT TTT CGT TTG ATG TAA CAA 1800
AAT AAG TCT AGA AAC TTA TTA AAC CCT TTA AAA GTA ATC TGA ATA CAG TTA AAC GTT ATT 1860

AAA CCT AAA GCC GTT ATT TTT TTG TTT TAA GGT GTA GTC TGG AAG AAT TCA TCT ACA AAT 1920
                                    P   T   T   Q   F   F   E   D   V   F   E

TCT AAA ATT ATA GCA TAA GTT AAA AGA CCG ATT GGT AAT AAA TTA CCC AAT AAA ACA AAA 1980
  L   I   I   A   Y   T   L   L   G   I   P   L   L   N   G   L   L   V   F   I

ATT CTT TTG GTT AAT TTC TCT TTA CTG AAC ATT GAA ATG ACG CTT AAA GGA ACA GCA AAT 2040
  R   K   T   L   K   E   K   S   F   M   S   I   V   S   L   P   V   A   F   L

AAG GAA GCA AAA AAT GTA ACT GAT ATA AAA ACC TTT AAA ATA GAT ACT ATA TTT TCA TTT 2100
  S   A   F   F   T   V   S   I   F   V   K   L   I   S   V   I   N   E   N   S

GAA GTC ATT GCT AAC ATA TAT AAG GAT ATT GGT AGA ATT ATA TTT GAA GTT AAT GCT GTT 2160
  T   M   A   L   M

AAG AAT GAA ATT GTA GTA AAC ACC CTC AAA TTT TCC CTT TTC AAA AAA TCC CCT CCA TTT 2220
ACT AAG AAT AAT GAT AGT TCA TTT GTA ACT TTG TTA CAA TGG ACA TTA AGA GAA TAT TTA 2280

                                        M   V   W   N   R   T   E   N   G   G
ACC CTG AGA ACA TGT AGG AAG AAA GCT TTA AAT GGT GTG GAA TAG AAC TGA AAA TGG AGG 2340

 T   N   L   L   T   T   L   I   F   D   I   D   G   T   I   L   D   T   E   K
GAC TAA TTT GCT TAC AAC TCT TAT CTT CGA TAT TGA TGG TAC AAT TTT AGA TAC AGA AAA 2400

 A   I   L   K   S   L   Q   R   I   L   K   E   E   L   S   T   D   Y   T   L
AGC AAT ATT AAA ATC GTT ACA AAG GAT TTT GAA AGA AGA ATT AAG TAC AGA TTA TAC ATT 2460

  M   R   G   I   C   I   G   Y   P   G   K   E   A   L   K   K   L   N   V   P
  D   A   W   N   L   H   W   V   S   W   E   G   S   F   K   K   I   E   C   A
AGA TGC GTG GAA TTT GCA TTG GGT ATC CTG GGA AGG AAG CTT TAA AAA AAT TGA ATG TGC 2520

  N   I   D   V   I   H   P   K   W   S   K   T   V   L   D   Y   N   H   E   V
CTA ATA TAG ACG TTA TTC ATC CAA AAT GGA GTA AAA CTG TAT TAG ATT ATA ATC ATG AAG 2580

  S   V   F   K   S   L   E   D   I   I   G   M   L   S   L   S   N   I   K   L
TCT CCG TGT TCA AGT CAC TTG AAG ATA TCA TTG GTA TGT TGT CAC TAA GTA ATA TAA AAC 2640

  G   I   V   T   S   E   T   K   Q   E   L   I   D   E   F   E   P   F   G   L
TAG GTA TTG TGA CGT CGG AAA CTA AAC AGG AAT TAA TAG ACG AAT TTG AGC CAT TTG GCT 2700

  S   S   Y   F   E   H   A   C   D   T   D   K   H   K   P   H   P   E   P   L
TAA GTA GTT ATT TTG AAC ATG CAT GTG ATA CGG ATA AGC ATA AAC CAC ACC CTG AAC CTT 2760

  L   A   C   L   K   G   L   D   V   P   C   H   E   A   I   Y   I   G   D   S
TAT TGG CCT GCT TGA AAG GAT TAG ACG TCC CTT GTC ACG AGG CAA TTT ACA TTG GAG ATT 2820

  I   Y   D   M   Q   C   T   K   S   A   G   V   K   F   A   L   A   S   W   G
CTA TAT ATG ATA TGC AAT GTA CAA AAA GTG CTG GTG TGA AAT TTG CTC TTG CTT CGT GGG 2880

  S   K   T   N   D   A   F   K   D   T   D   Y   I   L   E   E   P   K   D   I
GAT CAA AAA CAA ATG ATG CTT TTA AAG ACA CTG ATT ATA TTT TAG AAG AAC CAA AAG ATA 2940

  L   R   L   I   *
TTT TAA GAT TAA TAT AAA AGG TTA ATG CTT TAC TTT AAC TAG ATT AGA TAT TGC AAA ACC 3000
```

Figure 8.3 continued.
Nucleotide sequence of the 9800 bp of Bt ssp. *fukuokaensis* DNA upstream of *am1*. The sequence of *orfF* is shown from 2311 to 2955.

```
M   A   K   Y   T
AAA TAT TAA GTA ACA TTA AAT TAA ATA AAA AAG GAA GTG AAA AAT ATG GCT AAA TAC ACT 3060

E   E   Q   F   I   Y   F   T   T   K   L   N   Q   L   E   S   V   K   E   N
GAA GAG CAA TTT ATA TAT TTT ACA ACT AAA TTG AAC CAA TTG GAG TCT GTA AAA GAA AAC 3120

Q   E   T   Y   W   Y   E   Y   T   K   L   Q   D   W   L   Q   E   Q   Q   L
CAG GAA ACA TAT TGG TAT GAA TAT ACA AAG TTA CAA GAC TGG TTG CAA GAA CAG CAA CTA 3180

I   T   S   F   I   R   W   G   E   K   R   M   K   K   *
ATT ACC AGT TTT ATT AGG TGG GGT GAA AAG AGG ATG AAG AAA TAG TTC TTA ATT TAT CCT 3240
TTAGGGGTAC CGCAACGATT CGTTACTTTT GAAGCCACTG AGGAAAAATA TACTTTATAA 3300
TCCTATTTTT ATAGATTTTT GTATTCGGTA GTAACAAAAA TGGAAGTTTA TGTTACAAAT 3360
AGCACCACTA AGGCTGAATC CTAAATTCAT CTATAAGTGA TATGGAGTAT TTTGTATTAA 3420
CACCTCTTTT CGGAGGGCGA AATTATCCGA ATTAGGAATA AATAGGGTGT TTCTGCTCAA 3480
TTCCCCTTTT TCGAAGATGT GCAAAATACC TACATTAGAT TTTTAAATTT GGGAGGGAAA 3540

                                                M   N   K   Q   G   I   S   N
ATA GAG TGG GGC ATG AAA TAA GTT ATT TGA TTT CAG AAA TGA ATA AGC AAG GGA TTT CAA 3600

  E   H   I   E   S   M   K   E   R   I   M   Q   S   S   K   N   Y   E   E   H
ATG AAC ATA TAG AGT CAA TGA AAG AAC GGA TAA TGC AAA GCT CAA AAA ATT ATG AAT TTC 3660

  Y   D   K   V   N   E   N   K   E   H   S   L   V   D   V   N   K   I   K   G
ATT ACG ATA AGG TAA ATG AAA ATA AAG AGC ATA GTC TTG TTG ATG TAA ATA AGA TTA AAG 3720

  I   D   T   G   W   C   T   A   N   R   S   I   Y   E   L   F   F   S   V   L
GAA TTG ATA CAG GAT GGT GTA CAG CAA ATA GAT CAA TTT ATG AAC TGT TTT TTT CTG TAC 3780

  S   E   F   P   T   N   N   Q   R   L   D   G   R   K   I   G   Q   N   I   N
TAA GTG AAT TTC CAA CAA ATA ATC AAA GGC TTG ATG GTC GAA AGA TAG GCC AGA ATA TAA 3840

  N   L   I   E   N   G   I   Q   S   Q   Y   D   F   Y   A   N   K   M   G   N
ATA ATC TTA TAG AAA ATG GGA TTC AAT CTC AAT ATG ATT TTT ATG CAA ATA AAA TGG GAA 3900

  E   S   I   V   D   F   P   R   F   V   H   Y   I   D   D   D   I   Y   F   T
ACG AAT CAA TAG TTG ACT TTC CTA GGT TTG TTC ATT ACA TAG ATG ACG ATA TTT ATT TTA 3960

  V   S   D   A   T   H   R   T   V   S   A   I   M   F   D   A   P   K   M   M
CGG TTT CTG ATG CAA CGC ATA GAA CAG TGA GTG CTA TAA TGT TTG ATG CTC CTA AGA TGA 4020

  G   Y   V   T   T   Y   K   K   N   I   V   K   Y   N   I   Y   L   T   H   K
TGG GAT ATG TCA CTA CTT ATA AGA AAA ATA TTG TTA AAT ACA ATA TTT ATT TAA CAC ATA 4080

  D   S   V   Y   K   W   K   T   F   L   N   T   E   F   N   R   I   T   I   T
AAG ATT CTG TAT ATA AGT GGA AAA CTT TTT TAA ATA CAG AAT TTA ATC GCA TCA CAA TAA 4140

  R   T   P   T   D   K   Y   S   Y   D   G   E   Y   E   I   R   L   K   E   F
CCC GTA CTC CTA CAG ATA AGT ATA GTT ATG ATG GGG AAT ATG AGA TTC GAT TAA AAG AAT 4200

  P   K   K   Q   L   Y   F   K   F   S   T   P   I   V   T   E   T   Y   E   N
TTC CTA AAA AAC AAT TAT ATT TTA AAT TTA GTA CTC CAA TCG TTA CAG AAA CTT ATG AAA 4260

  I   I   D   N   T   L   L   K   K   R   E   P   Y   R   R   I   N   T   K   V
ATA TAA TTG ATA ACA CTT TGT TGA AGA AGA GAG AAC CAT ATC GGA GAA TTA ATA CAA AAG 4320

  A   A   D   R   K   T   F   I   *
TTG CAG CAG ATA GAA AAA CGT TTA TCT GAA AGT TCG AAC GAT ATT GGG GTT TTG CCG TTA 4380
```

Figure 8.3 continued.
Nucleotide sequence of the 9800 bp of Bt ssp. *fukuokaensis* DNA upstream of *am1*. The sequence of *orfG* is shown from 3578 to 4344.

```
TTACTACGTT TCATGAAGAA AATAAATTTC AACTTGGGGT ACTACCTTTT CTATGCGTTA 4440
TATAAGAAAG AAGTGTATTT GGATATAGAT GATCCATCTG ATGTTACAAG ATGGAAAATA 4500
CAAAAAATTA TTCGCAATGA AGTTATGATT GACAATTTTC CGAGAAAATG AAGTTTTTAA 4560
TTAGTTTTAC AGAATACACA AAATGTTAAA CTCAAAAAAA CAGTGATTTC ATGTGGGGTA 4620
TGGCAACGAT TACTTCTTTT TTGAGTCACT CATCTTTTTT TTAACAAAAA GGCTTAAAGT 4680
CAATAATGTT ACGAATTATT TTTGATACAG TATTTTGTTA CAGCTAATTC AACAGTTAAG 4740
AAATTTATAT TTTTAAATTT TCCTCAGTGG CTCCAAAAGT AACGAATCGT TGCCGTACCC 4800
TCTTTAGAAG TATTCCTTTT ATCATCAGGA TGAAAATGTA TTTAATTAAT AAATAGGGTA 4860
AGTGAGGCAT CCAGAGGGAT GCTTTTTCTT TCGGCATAGA AATCCCCTTG GTGACTGTCT 4920
TTTAAATATA ATTAATAAGT ACGGTGTGTG AAATTTTATA AAGTGTGTAG AGATGAGATT 4980
ATATTTAATA CAATTAGGGA AGAATGGCAT ATAAGATGAT GGATTAATAC ATTTAATTTT 5040
ATAAAAAGGG ATGAGTGTTA TGAACGCAGT TTCTAATACT TATCATATTT TGTATTATTA 5100

          M   P   F   P   S   I   T   L   N   Y   I   N   E   L   D   I   L   M
CTG TTT ACA TGC CTT TTC CAT CCA TCA CGC TTA ATT ATA TAA ATG AAC TAG ATA TAC TGA 5160

  Q   Y   I   I   K   S   V   Y   F   S   F   P   I   S   L   V   F   L   F   I
TGC AAT ACA TAA TTA AAT CTG TTT ATT TTT CTT TTC CTA TTT CAT TAG TTT TCC TAT TTA 5220

  M   K   D   Q   R   F   L   F   F   K   K   I   S   F   Q   Q   L   K   I   S
TAA TGA AGG ATC AAA GAT TTT TGT TTT TTA AAA AAA TCT CTT TCC AGC AAC TAA AAA TTT 5280

  F   F   V   L   S   L   F   M   L   L   C   C   S   F   *
                                        M   L   F   F   L   A   V   L   I   S
CTT TTT TTG TGC TTT CAT TAT TTA TGC TAT TAT GTT GTT CTT TTT AGC CGT ACT AAT TTC 5340

 F   T   T   F   K   G   T   S   N   P   I   G   N   E   G   I   T   N   L   L
TTT TAC AAC ATT CAA AGG AAC ATC TAA TCC TAT TGG GAA TGA AGG AAT TAC TAA TTT GCT 5400

 Y   K   T   P   G   I   A   I   Q   L   I   G   E   N   I   M   L   V   S   I
TTA CAA AAC ACC TGG GAT TGC AAT TCA ATT AAT CGG AGA GAA CAT TAT GTT GGT ATC TAT 5460

 L   F   F   W   H   K   I   I   R   S   F   I   I   S   P   I   S   S   I   T
ATT GTT CTT TTG GCA TAA GAT CAT ACG CTC CTT CAT AAT AAG TCC CAT ATC TTC AAT TAC 5520

 S   S   L   I   L   S   G   S   S   F   G   L   L   H   L   S   T   Y   N   Y
ATC CTC GTT AAT ATT GTC TGG AAG CTC ATT TGG GTT ACT CCA TCT CTC TAC GTA TAA CTA 5580

 N   W   I   Q   C   L   T   I   I   G   I   P   A   I   A   Q   M   I   F   F
TAA TTG GAT ACA ATG TTT AAC TAT CAT TGG GAT ACC AGC TAT CGC TCA AAT GAT TTT CTT 5640

 L   I   F   R   N   I   H   M   G   Y   I   L   H   F   N   Y   N   F   I   I
TCT AAT ATT TAG AAA TAT TCA TAT GGG ATA CAT ACT TCA CTT CAA TTA CAA TTT TAT CAT 5700

 V   L   F   S   Y   I   A   S   I   *
TGT ACT TTT TAG CTA TAT AGC TTC TAT TTA AAA TAT TTT ACC ACT GAA AAT AAA AGA AAA 5760

GGT ACC CCT GCT GCA AGG TAC CTT TTC TAA GGG GTA TAA CTC ATT AGC TAA GCG GCC GTT 5820
GGT CAA GTA TAC CGC CTA GAC TAT TAC ATG AGT TAT GTT TTT ATT ATA CAT TGA TCA TAT 5880

CAA AAA TAT TAT CAT TAG TGA AAT TAA CTT CTA TTA TTT TAT AAT TTT AAA GCT GCC ACC 5940
                                        K   I   I   K   F   S   G   G

GTA ATA ATA ATT TTC TCC AAG TTT CAT AAC ACC TAT TAC ATC AGT ACC AGT ATC ATA AAT 6000
Y   Y   Y   N   E   G   L   K   M   V   G   I   V   D   T   G   T   D   Y   I

TGT TTT ATA TCT TTG CAT ATT TAC TAC TTC ATT CTC TTC TTC ACT TAA TCG ATA TAT CTT 6060
T   K   Y   R   Q   M   N   V   V   E   N   E   E   E   S   L   R   Y   I   K
```

Figure 8.3 continued.
Nucleotide sequence of the 9800 bp of Bt ssp. *fukuokaensis* DNA upstream of *am1*.

```
TAC AAT ACC TTT ATC TTT ACC GCC TGT ATA AAT TTT AAT GTC CAT AGT ATC TCC ATG CTT 6120
V   I   G   K   D   K   G   G   T   Y   I   K   I   D   M   T   D   G   H   K

TAA ATC ACC AAC AGC TAA TTC TTG TAA GAA GTC TGC ATT ATG ATG AAG TAA AAA ATA GTT 6180
L   D   G     V   A   L   E   Q   L   F   D   A   N   H   H   L   L   F   Y   N

TTG CCA TTG CCA GTC ATC TAA AAC CGG TGA TGT TTT TTT GTA TAA ATT GAG ATT CAG CAG 6240
Q   W   Q   W   D   D   L   V   P   S   T   K   K   Y   L   N   L   N   L   L

AAG CAG GTG TAA TGC TTC CTA CAG TAT AGG AAA ATA CAC TTC CTC CAA TAG TTA ATG CTC 6300
L   L   H   L   A   E

CAATAATAGC TACTTTTCCT AAAAGTTTCA TTTTTAAGCC TCCTAGTTTA TTAAGAACTA 6360
TAAAATCTTA CTCATGAAAA ACAGCTATCA AATCATTAGA TGGTAATATT TTCTATACAA 6420
ATTTTAACAT ATGAAGATTT TTCTGCCAAT CAATTGATAA AAATATGGAA TAAAATAAAG 6480
TTGAATTATA AACACTCTGA AAATCCTTAG TAACTTAACA CTAAAAAGAA TCCATAAAAA 6540

AAA TAG ATT AAA CAA ATT TGA TTC TGA CTC GTG AAA TTG CTG CTA GCC GCA CCT TTT AAA 6600
                                                    *   G   C   R   K   F

AGA AGC GTT ATA ACA AAA AAC ACA CAT TAT TGA TAC AAT CAT AAG CTT AAT GGG AAT ACT 6660
S   A   N   Y   C   F   V   C   M   I   S   V   I   M   L   K   I   P   I   S

AGT ATC ATT AGC GGG AAA GGT TAA TGG TAA TAT ATT AGC TAT TAT AGC CGC TCC AGC TGC 6720
T   D   N   A   P   F   T   L   P   L   I   N   A   I   I   A   A   G   A   A

TAT AGA TGT TCC GCC AGT AAT TGG AGC GGC TAA CAA TCC TGC AAC AGC TGC AGC GAT ACC 6780
I   S   T     G   G   T   I   P   A   A   L   L   G   A   V   A   A   A   I   G

GAA AAC TGA AAC AAG AGC ATT TTG CCA GGT AAG GTC TAA AGG TTC TAT AGA ATT AAG GTT 6840
F   V   S   V   L   A   N   Q   W   T   L   D   L   P   E   I   S   N   L   N

ATT ATT TGT ACA CAT ATT TAG CCA GTC TTT ATA ATT CGT ATT TTG TAT AGA TGT TTG TGG 6900
N   N   T   C   M   N   L   W   D   K   Y   N   T   N   Q   I   S   T   Q   P

ATC TTT TAC TAG TGG ATA CCT TGG ATA ACG ATT AGA CAT ATT AGA GTT GTT TCG TGA AGC 6960
D   K   V   L   P   Y   R   P   Y   R   N   S   M   N   S   N   N   R   S   A

ATC CAA TAT TTC ATA TTC ATT TTT ATT TTC ATA TGA ATT CAT CTT TTT CAC TCT TTA TAT 7020
D   L   I   E   Y   E   N   K   N   E   Y   S   N   M

AGT CCT TAT TCT TAA GAC ACC TAA CTA ATA TAT GCC AGG TGT CTT TTT ATA TAT TCG TAA 7080
                            *   Y   I   G   P   T   K   K   Y   I   R   L

TGC CCC TGT GTA TTC CGG TAA TAA TGT TTC ACC AAA TAA CCC CAT TAT TTT TGG TGA AAG 7140
A   G   T   Y   E   P   L   L   T   E   G   F   L   G   M   I   K   P   S   L

TAT TGA TTT TTT AAA CCT CTC TTT ATG TAT TGC ATG TAT TTT TTT TAA AGA TTC AGA TTT 7200
I   S   K   K   F   G   R   K   I   Y   Q   M   Y   K   K   L   S   E   S   K

CTG TAG AAC GAT TTT TTC TTT GTT GTT TTT AGG AGT GAG GAA AGG TGA ACT CAC TTT TAA 7260
Q   L   V   I   K   E   K   N   N   K   P   T   L   F   P   S   S   V   K   L

ATA AAC AAA ATT TAC ACT TGG GTG CTT TTG CTG GTT CTG GTG CAA ATA ATA AAC AAA GGC 7320
Y   V   F   N   V   S   P   H   K   Q   Q   N   Q   H   L   Y   Y   V   F   A
```

Figure 8.3 continued.
Nucleotide sequence of the 9800 bp of Bt ssp. *fukuokaensis* DNA upstream of *am1*. The sequence of the highly truncated *cry* gene AM3 is shown from 7002 to 6585.

```
TAC AAA GGG CAT AGA TTC CTA ATG GAA TCC ACT CTT TAA GCA GGT AAT CTA AAT AAC TGA 7380
V   F   P   M                                   *   A   P   L   R   F   L   Q   H

TGA ATA CAT CTA TTC TGA TTT TTG AGA GAC TGG ACC CTT AAC TTA AGT TGA CCT TTT TTG 7440
  I   C   R   N   Q   N   K   L   S   Q   V   R   L   K   L   Q   G   K   K   I

ATC ATA TGC ACA ACT TCT ATT CCA GCA ATC ATT TTT GTA GCT GTG CGC AAT GAT TTT AAT 7500
  M   H   V   V   E   I   G   A   I   M   K   T   A   T   R   L   S   K   L   G

CCA AGC ATA TTC CGG ATT CGT TTT TTG ATA AAG CGA TGG TCC TGT TCA ATA ATG GTA TTT 7560
                                                        *   Y   H   Y   K
  L   M   N   R   I   R   K   K   I   F   R   H   D   Q   E   I   I   T   N   L

AAA TAT TTT TTC ACA CGA AGT GGA GTA TCG TGT GGA AGG CGC TTC TCC TCT TTT AAT TTC 7620
F   I   K   E   C   S   T   S   Y   R   T   S   P   A   E   G   R   K   I   E
  Y   K   K   V   R   L   P   T   D   H   P   L   R   K   E   E   K   L   K   Q

TGT ATC ATC ACC ATC GAC CGT AAA TAT ACG AGG TTT TGT GAC ATG ACA AGA AGC CAA GGC 7680
T   D   D   G   D   V   T   F   I   R   P   K   T   V   H   C   S   A   L   A
  I   M   V   M   S   R   L   Y   V   L   N   Q   S   M
      ##############
TTT CTT TAG AAA ACG CTT GGC AGC CTT CGC ATC TCG TTT ACT ACT CAA ATA AAA ATC AAG 7740
K   K   L   F   R   K   A   A   K   A   D   R   K   S   S   L   Y   F   D   L

CGT GTT TCC TTC GGA ATC AAC AGC ACG ATA TAA ATA CAT GTT CTC ACC TTT AAT TTT GAG 7800
T   N   G   E   S   D   V   A   R   Y   L   Y   M   N   E   G     K   I   K   L

GTA TGT TTC ATC CAC TCT CCA AGA ATC ATT CGT TGG TTT CAA ATG TTT TCG AAT CCG TTC 7860
Y   T   E   D   V   R   W   S   D   N   T   P   K   L   H   K   R   I   R   E

ATT TAA TTC TGG CCC ATA TTG GTG AAC CCA GCG CAT AAT GGT TGT ATA GGA CAT AGA TAA 7920
N   L   E   P   G   Y   Q   H   V   W   R   M   I   T   T   Y   S   M   S   L

TCC TCG TTC TTC CGT CAT TTC GAC CAA ATC ACC AAA GCT TAG GCT GTA CCG CAG GTA CCA 7980
G   R   E   E   T   M   E   V   L   D   G   F   S   L   S   Y   R   L   Y   W

TCG TAC AGT TAA CAA AAT AAT GTC TGG TTG ATA ATG TTT CCA TTT GAA TAT ATT TTC TTT 8040
R   V   T   L   L   I   I   D   P   Q   Y   H   K   W   K   F   I   N   E   K

TTC CAT CCT GAT CAC GTA CCT TTT TTT AGT AGC GGT ACC AGT ATG GCC GAA TTA GAA AAA 8100
E   M

AAT TTT TCA AGT TTT TTA TTT TTT TGC ACC AGA ACC AGA ATA TCT CTT GTA TCA TGC TCA 8160

TCG AGT CAT TTT TTT AAT TTC AAT TTT ACA GAA GAA CCT ATT GAT GAA TAT TAT CAA TTA 8220

TTC ACT AAT ATC TTG GTA TTA GGT TCA AAA TAA GAA TTA CGA AAG GGT GCT TTT TGA TGA 8280

                                                            M   K   N   T
ATA ATT ACA GTA ATA TGA ATT ATA ACG GCG ATA AAG AAG AAC CTA GAT ATG AAG AAT ACC 8340

Q   G   M   R   M   I   I   L   G   V   Y   Q   N   I   Q   G   N   A   H   I
CAA GGT ATG AGG ATG ATC ATT TTG GGT GTA TAC CAA AAC ATC CAA GGC AAT GCA CAT ATT 8400

V   K   D   N   I   I   N   N   V   D   V   N   V   L   Q   H   V   L   Q     H
GTG AAG GAC AAT ATA ATC AAC AAT GTG GAT GTA AAT GTC CTC CAG CAT GTC CTC CAG CAT 8460
```

Figure 8.3 continued.
Nucleotide sequence of the 9800 bp of Bt ssp. *fukuokaensis* DNA upstream of *am1*. The two overlapping open reading frames encoding the putative IS*240*A insertion sequence (*T2*) run from 7665 to 7356 and from 8046 to 7548. The point at which the frame shift occurs is marked (#) and underlined, the IS sequence reads DGVM in this section. The sequence *orfI* runs from 8328 to 8991.

```
V   L   Q   D   H   K   G   L   Q   G   L   Q   G   L   K   D   I   P   G   P
GTC CTC CAG GAC CAC AAG GGC CTC CAG GGC CTC CAG GGC CTC AAG GAT ATC CCA GGG CCT 8520

Q   G   Y   P   G   P   P   G   P   L   E   V   P   I   S   G   V   F   F   S
CAA GGA TAT CCA GGA CCT CCA GGA CCA TTG GAG GTT CCA ATA AGT GGC GTG TTT TTT TCC 8580

S   T   E   N   G   T   G   S   V   V   V   P   G   N   H   N   Y   P   L   K
TCA ACG GAA AAT GGC ACA GGA TCT GTA GTT GTT CCA GGT AAT CAT AAT TAT CCT TTA AAA 8640

S   F   R   E   T   V   P   G   S   F   S   L   A   A   N   E   I   T   I   N
TCT TTC AGG GAA ACA GTA CCT GGA TCA TTT TCA TTA GCT GCT AAT GAA ATT ACA ATT AAT 8700

V   S   G   I   Y   L   V   D   G   Y   V   V   L   A   R   T   S   G   A   G
GTT TCA GGG ATT TAT CTA GTA GAT GGA TAT GTA GTT TTA GCA AGA ACA AGT GGT GCC GGT 8760

N   A   R   I   I   V   N   N   V   N   S   G   T   I   T   M   R   A   N   P
AAC GCT AGG ATT ATT GTA AAT AAT GTG AAT TCC GGC ACG ATT ACT ATG CGT GCA AAT CCT 8820

G   E   Q   T   I   C   N   G   Q   V   S   L   N   V   G   D   V   V   K   F
GGA GAA CAA ACA ATC TGT AAT GGT CAA GTT TCT TTA AAT GTT GGA GAT GTA GTG AAA TTT 8880

I   N   F   G   S   F   G   S   S   T   V   T   T   S   A   E   L   N   I   Q
ATT AAT TTT GGA TCT TTT GGT TCA TCA ACA GTA ACA ACA TCT GCA GAA TTA AAT ATT CAA 8940

A   N   Y   N   N   V   Q   V   R   L   T   R   L   G   P   L   T   *
GCT AAT TAT AAT AAT GTT CAA GTA AGA TTA ACA AGA CTA GGC CCT TTA ACA TAG ATA AGT 9000

GTGTTGGTTC TGTTGCAAAC CTTTAAATTC TCTACACTTA GCCTTTAAAA AAAGAAACAA 9060
CCTCTACTAT AATTGAAGTT GGATTTGCTT AACTTGATAG CGATGTGGTG ATATCCCTAA 9120
TTATGTTGGA TATATTAAAG TTTAGTAATT GCATTATGAA TTAAAACTAT TGAGTAAAAA 9180
ATCCGACTCA TAAGCTTTAA GTCGGATTTT TATTTATAAT TTTTTATAAA GTTTCCCTAA 9240
ATCTTTCTTT ATCTGAGATT TCAGTTGATG ATTTTCGTCT TCTAGTGTCT TAACTTTCTG 9300
TCTTAGTAGC TTCATTTTAT CTTTTAAAGT TTCTATAATT ACAGTTTTCT GTACACAGGT 9360
AAAATTCCCA TAATTCGGTT TTTCCTAGAA AAAATAAAAT GAATCCAATA CAGGTTCGGA 9420
TTTATTTTCT ACAATTTCTT TCAATGACGA ATAATTAAAT ATCTACAACA CTCTATGTAG 9480
TATCCCTCTC TTTTTTAGAG GGGGGGAAAA AACATCCAAG GGTGAATTTC GTTCATATAA 9540
AAGTGAATGA CTCTTTTCGC ACCCTTAAAA ACAACAAAGA AAAAATCGTT CTATAGAAAT 9600
CTGAAACTTC AAAAAATTAT ATGCAATACA TAAAGAAAAG GTTTAAAAAA TCAATATTTT 9660
TACCAAAAAT AATGGGTTTA TTTGTAGAAA CATTGTCACA GGAATACATT GGGGCACTAC 9720
GAATATATAA CAAGACACCT AGCATATACT GTTTGGGTAT CTAAAAATAA GGACCACATA 9880

                  M   N   S   Y   Q   N   K   N   E   Y   E   I   L   D   A   S
AGG GAG TGA AAA ATA TGA ATT CAT ACC AAA ATA AAA ATG AAT ATG AAA TAT TGG ATG CTT 9940
```

Figure 8.3 continued.
Nucleotide sequence of the 9800 bp of Bt ssp. *fukuokaensis* DNA upstream of *am1*. The sequence *orfI* runs from 8328 to 8991. The start of *am1* is shown from position 9894.

Over this short region, AM3 was found to share 41.2% amino acid sequence similarity to AM1. The percentage similarities to the other Cry4 and Cry10 types (Figure 8.4a) indicate that AM1 and AM3 may have diverged more recently than AM1 and the other protein sequences analysed. The alignment in Figure 8.4b shows the high degree of similarity between these sequences. The discovery of yet another Cry4-type sequence, albeit a highly truncated one, is a further example of the dissemination of this increasingly diverse family of proteins. The theory that this dissemination is as a result of transposase activity is supported by the discovery of a second putative insertion sequence (T2) encoded by two overlapping open reading frames *orfX* and *orfY* upstream of *am1* (shown between nucleotides 7665-7356 and 8046-7548 respectively).

Alignment of *orfX* and *orfY* with IS240A (the sequence found to have the highest degree of similarity to *orfX* and *orfY*) indicates that a single frame shift mutation (marked in Figure 8.3) prevents the putative insertion sequence from being encoded as one continuous open reading frame. *orfX* and *orfY* are orientated in the same direction as the previously identified downstream insertion sequence *orfA* and may form a novel type II transposase with the putative integrase/recombinase *orfD*. Transposable elements in Bt are further discussed along with a more complete analysis of both putative insertion sequences and *orfD* in Section 8.5. Figure 8.5 represents the sequence data from Figure 8.3 in graphic form. The location of open reading frames showing significant similarity to previously identified protein sequences are indicated.

Percentage Similarity

	AM3	AM1	Cry4A	Cry4B	Cry10A	
AM3		41.2	30.4	19.6	31.1	AM3
AM1	53.8		29.5	28.1	32.7	AM1
Cry4A	67.4	66.2		24.4	44.2	Cry4A
Cry4B	71.7	62.7	68.6		27.6	Cry4B
Cry10A	66.4	65.5	54.9	67.8		Cry10A
	AM3	AM1	Cry4A	Cry4B	Cry10A	

Percentage Divergence

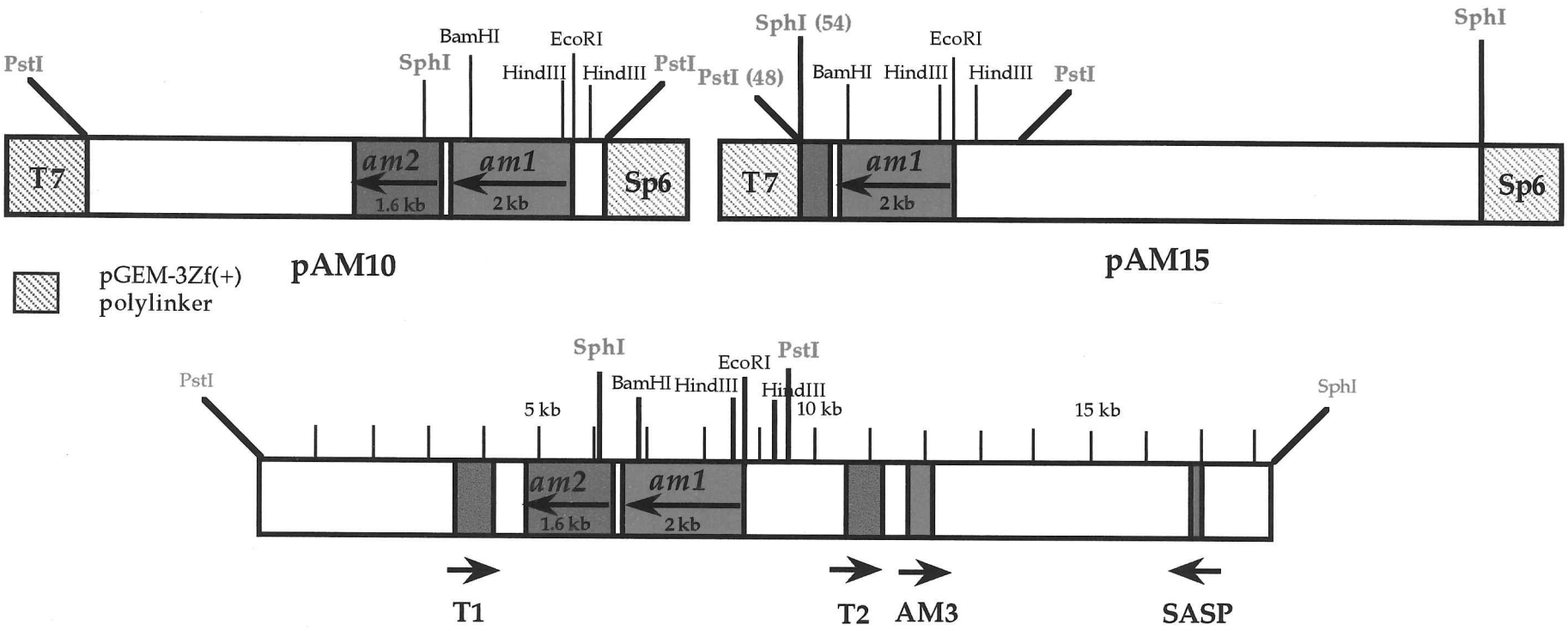
Figure 8.4a Protein sequence relationships using a PAM250 residue weight table.
Only the regions of AM1, Cry4A, Cry4B and Cry10A corresponding to the truncated AM3 are used in this alignment. The table represents the initial pair-wise matrix used in the first process of the multiple alignment. The percentage similarity between individual pairs can be determined from the upper right area in the table. Percentage divergence between sequences is shown in the lower left portion of the table and is calculated by comparing sequence pairs in relation to the phylogenetic tree (not shown).

```
AM3       MNSYENKNEYEILDASRNNSNMSNRYPRYPLVKDPQTSIQNTNYKDWLNMCTNNNL
AM1       MNSYQNKNEYEILDASRNNSNMSNRYPRYPLANDPQACMQNTNYKDWLATCNGTTV
Cry4A     MNPYQNKNEYETLNASQKKLNISNRYTRYPIENSPKQLLQSTNYKDWLNMCQQNQQ
Cry10A    MNPYQNKNEYEIFNAPSKGFSKSNRYSRYPIENKPNQPLKSTNYKDWLNNCQDNQQ
          ** * ********** *  * ** * ****   *    ********  *  *

AM3       NSIEPLDLTWQNALVSVFGIAAAVAGLLAAPITGGTSIAAGAAIIANILPLTFPANDTSI
AM1       PFSNP------SQLLKVGGIV--VQRVLGGIATFFPGIGPLVPFLTFFASLLWPSGSSGN
Cry4A     YG---GDFETFIDSGELSAYTIVVGTVLTG-FGFTTPLGLALIGFGTLIPVLFPAQDQS-
Cry10A    YGNNAGNFASSETIVGVSAGIIVVGTMLGA-FA-APVLAAGIISFGTLLPIFWQGSDPA-
                        *     *   *                        *   *

AM3       P-I--KLM------IVSIMCVFCYNASFKRCG
AM1       D-IWEKLMKEVADLIQQELTTYTIDKATAELG
Cry4A     N-TWSDFITQTKNIIKKEIASTYISNANKILN
Cry10A    N-VWQDLLNIGGRPIQ-EIDKNIINVLTSIVT
                      *
```

Figure 8.4b Clustal alignment of the amino acid sequence of AM3 with the corresponding regions of AM1, Cry4A and Cry10A.
Conserved or semi-conserved residues are shown in bold and marked with an asterisk. Residues that are conserved between AM3 and all but one other sequence are marked with an asterisk in plain text.



T1 - Novel Transposase from the IS6 family (*orfA*).
58% similarity to IS*214*-II from *Lactobacillus lactis*. The closest Bt relatives are IS*240*A (Delecluse *et al.*, 1989) from Bti (38% similarity), and 74% similarity to a newly identified transposable element from the related strain Bt ssp. *fukuokaensis* 84-I (Dunn & Ellar, 1997)

T2 - Novel transposase (*orfX*, *orfY*).
81% similarity to IS*240*A from Bti
The open reading frame is disrupted by a single frame shift.

AM3 - Open reading frame encoding the first 139 amino acids of a *cry* gene with 41% similarity to AM1 over this region.
N-terminal sequence :- MNSY **E** NKNEYE
(difference from AM1 marked in black).

SASP - A complete open reading frame encoding a 74 amino acid protein with homology to an α/β-type **small acid soluable spore protein**. The greatest similarity being 60.3% to the C5 protein of *B. megaterium*.

Figure 8.5 Genetic map of region surrounding *am1*. Constructs pAM10 and pAM15 are shown above a diagrammatic representation of the region of 17A DNA that they contain. The arrows indicate the transcriptional direction of the above genes.

8.5 Transposable Elements

8.5.1 Classification of transposable elements

Existing as discrete mobile sequences in the genome, transposable elements or transposons are able to transport themselves to other locations within that genome. Their presence was first identified in the form of spontaneous insertions into bacterial operons that prevented transcription and/or translation of the gene into which it was inserted. The ability of such transposable elements to mobilise other regions of DNA with which they are associated provides a potent force for change within both prokaryotic and eukaryotic genomes. They are therefore interesting, not just for the mechanisms involved in the manipulation of DNA, but also as has been briefly discussed above, for the evolutionary consequences of their mobility.

The characteristic of a transposon is that it does not exist as an independent form of the element but moves directly from one site in the genome to another. Unlike most other processes involved in genome restructuring, transposition does not rely on any relationship between the sequences at the donor and recipient sites. Transposons are restricted to moving themselves, and sometimes additional sequences of the genome, to new sites elsewhere in the same genome; they are therefore an internal counterpart to the vectors that can transport sequences from one genome to another.

Each bacterial transposon carries a gene, or genes, that codes for the enzyme activities required for its own transposition, although it may also require ancillary functions of the genome in which it resides (such as DNA polymerase or DNA gyrase). The transposable element is believed to constitute 'selfish DNA' concerned only with its own propagation and is considered very much as an independent entity that resides in the genome. Such a relationship between the transposon and the genome can be likened to that of a parasite with its host. The propagation of an element by transposition may cause harm if a transposition event inactivates a necessary gene, or if the number of transposons becomes a burden on cellular systems. However, it may also confer a selective advantage as is discussed later in this section.

Kleckner (1981) sub-divided transposable genetic elements into three major classes based on their mechanistic differences, structural organisation and sequence homology. Class III comprise the transposing bacteriophages Mu and D108 and are not discussed.

Class I are insertion sequences (IS) and composite elements formed from them. Normal constituents of bacterial chromosomes and plasmid IS elements exist as autonomous units, each of which encodes the proteins necessary for its own transposition. They are all short (approximately 750-1500 bp) and end in inverted terminal repeats that serve as a substrate for a 30 to 35-kDa polypeptide, the 'transposase' essential for transposition activity (Iida *et al.*, 1983). This transposase is encoded by the major coding

capacity of an IS element. The element may also contain shorter open reading frames, in some cases overlapping the transposase gene (e.g. IS*3*, Timmerman & Tu, 1985; IS*150*, Schwartz *et al.*, 1988). IS elements can be grouped into families based on the homology of the transposase sequences and the length and composition of the terminal inverted repeats. The two copies of the inverted repeats are short (9-23 bp) and are usually closely related rather than identical.

When an IS element transposes, a sequence of host DNA at the site of insertion is duplicated, therefore the IS DNA is always flanked by very short direct repeats, the most common length for which is 9 bp. The sequence of the direct repeat varies among individual transposition events undertaken by a transposon but the length is constant for any particular IS element. An IS element therefore displays a characteristic structure with inverted terminal repeats at its ends, while the adjacent ends of the flanking host DNA are identified by the short direct repeats.

Class II are much longer elements (5 kb) and encode accessory determinants not strictly required for transposition. A number of Class II transposases (Tn) are themselves flanked by insertion sequences, for example Tn*9* and Tn*1618* are flanked by copies of IS*1* (MacHattie & Jackowski, 1977; So *et al.*, 1979). The Class II transposases share similarity in their structure and function to Tn*3*, the best characterised transposon in this group (reviewed by Heffron, 1983). The three main features that these elements have in common apart from their terminal DNA repeats are;-

(a) They encode a high molecular weight protein essential for transposition (TnpA, about 120-kDa).

(b) In addition, they encode a smaller polypeptide (TnpR) required for the site-specific recombination and resolution of the cointegrate formed as an intermediate in the transposition pathway.

(c) Related to TnpR is the site of this resolution (res).

These elements are true transposons whereas composite transposons are formed from two or more copies of a Class I insertion sequence flanking a gene or genes. The flanking IS elements may be identical but are often only closely related and may either be in the same or inverted orientation.

Transposition of Class II elements proceeds in two steps (Heffron, 1983); a transpositional replicative recombination (requiring the *tnp*A gene product and the terminal inverted repeats) followed by a site-specific recombination mediated by the resolvase protein (*tnp*R gene product). The first event leads to the fusion of donor and recipient replicons, with the two copies of the transposon in direct orientation at their junctions. This molecule is known as a cointegrate, which must be resolved at the next stage by a site-specific dsDNA exchange, resulting in the integration of a single transposon in the recipient DNA. All Class II transposons transpose through this two-step procedure in agreement

with the asymmetric recombination mechanism proposed by Shapiro (1979) and Arthur and Sherratt (1979).

Transposition of IS elements is thought to occur through a conservative mechanism. The inverted repeats are regarded as the site that transposases recognise, cut and rejoin to the target sequence. With the exception of IS*1* (Biel & Berg, 1984) and IS*903* (Weinert *et al.*, 1983) transposition by Class I elements is thought to proceed through a conservative mechanism since members of this group rarely form cointegrates and are incapable of resolving them when they are formed.

Both Class I and Class II transposable elements can be associated with genes that confer a selectable phenotype on the host. For example Tn*3* carries the β-lactamase gene conferring ampicillin resistance on the host bacterium (Heffron *et al.*, 1975). In a similar way, the association of both Class I and Class II transposable elements with the *cry* genes in *Bacillus thuringiensis* may be thought of as conferring the selectable advantage of entomotoxicity to the host bacterium. Figures 8.6 and 8.7 show the organisation of several Class I and Class II transposable elements from different Bt strains. To date three different Class I elements have been found to flank *cry* genes forming composite transposons: ISI*231* (Mahillion *et al.*, 1985), IS*232* (Menou *et al.*, 1990) and IS*240* (Delécluse *et al.*, 1989). An IS*150*-like element (Smith *et al.*, 1994) found upstream of *cry*1C (Smith *et al.*, 1994) was shown to be present in a single copy and so is thought not to be part of a composite transposon. Two Bt. Class II transposons have also been characterised, Tn*4430* (Lereclus *et al.*, 1984) and Tn*5401* (Baum, 1994). More recently two IS-like elements and a putative site-specific recombinase (*orfRec*) were identified upstream of *orf1* from Bt ssp. *fukuokaensis* 84-I (Dunn & Ellar, 1997). These Bt Class I and Class II elements will be discussed further in this section.

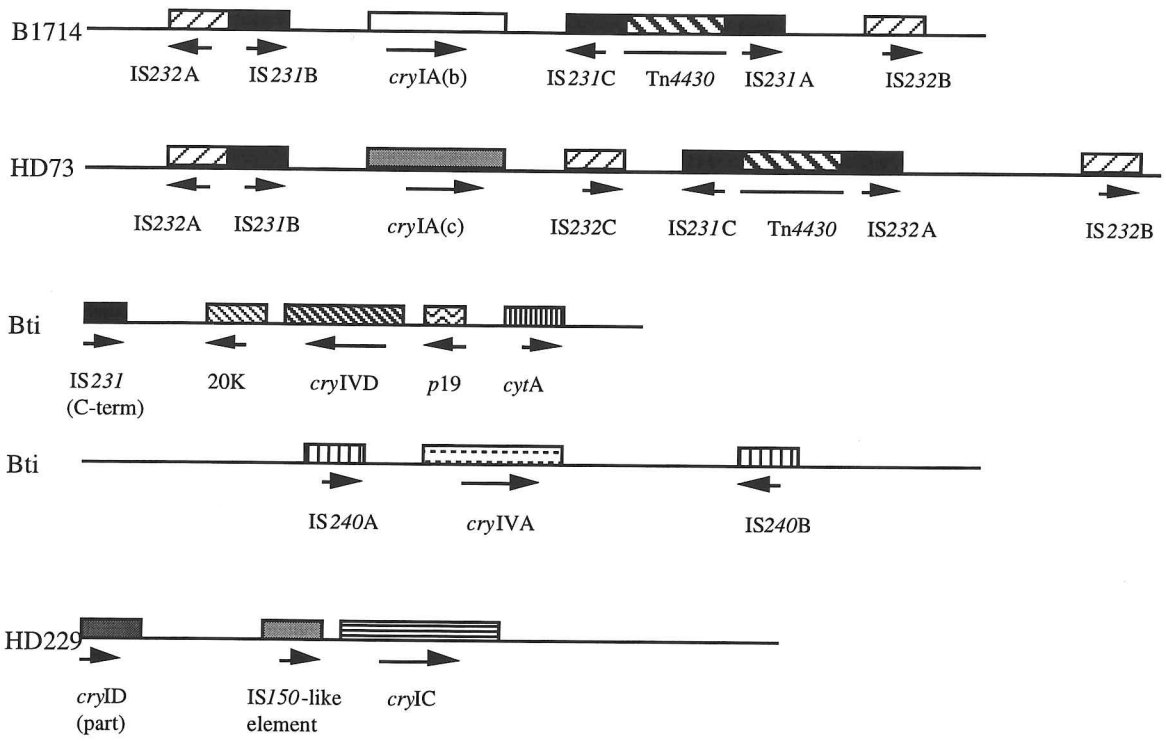


Figure 8.6 Structural organisation of transposable elements flanking δ-endotoxin genes resident on large plasmids of lepidopteran and dipteran active Bt strains (Menou *et al.*, 1990; Delécluse *et al.*, 1989; Bourgouin *et al.*, 1988; Adams *et al.*, 1989; Lereclus *et al.*, 1984; Mahillon *et al.*, 1987; Smith *et al.*, 1994).
Arrows indicate the relative orientation of the elements, and for *cry* genes the direction of transcription. Tn*4430* can be found in opposite orientations in different strains. Abbreviations: B1715; Bt ssp *berliner* 1715, HD73; Bt ssp *kurstaki* HD73, Bti; Bt ssp *israelensis*; HD229; Bt ssp *aizawai* HD229, IS; insertion sequence, Tn; transposon, C-term; C-terminal fragment.
Taken from Dunn (1996).

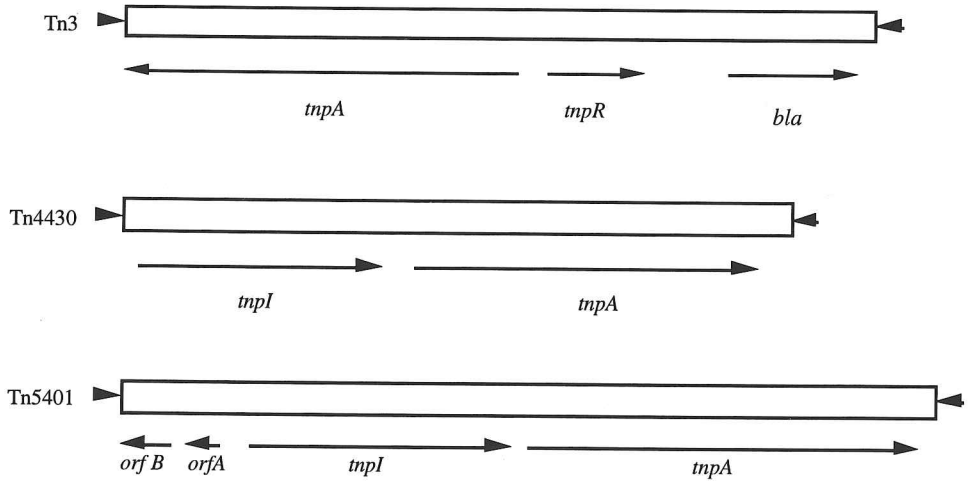


Figure 8.7 Structural organisation of Tn*3*, Tn*4430* and Tn*5401*.
Tn*3* contains 4957 bp and codes for three polypeptides; TnpA: a transposase required for transposition, TnpR: a site-specific recombinase, and *bla*: the β-lactamase gene conferring ampicillin resistance (Heffron *et al.*, 1979). The Bt transposon Tn*4430* contains 4149 bp and encodes two polypeptides; TnpA, related to TnpA from Tn*3*, and TnpI encoding the putative resolvase gene (Mahillon & Lereclus, 1988). Both Tn*3* and Tn*4430* are flanked by 38 bp inverted repeats. The Bt transposon Tn*5401* contains 4837 bp and codes for two major polypeptides; TnpA: a transposase related to TnpA from Tn3 and TnpI encoding a putative resolvase (Baum, 1994). It also encodes two smaller Orfs of unknown function (Orf1 is 85 residues and Orf2 is 74 residues). Unlike Tn*3* and Tn*4430* it is flanked by 53 bp inverted repeats.

8.5.2 Class I Mobile Elements in Bt

The existence of composite transposons bearing δ-endotoxin genes, was first suggested by the discovery of transposable elements IS*231*A (corresponding to IR*1750* (Kronstad & Whiteley, 1984) from Bt ssp. *kurstaki* HD73) flanking *cryI* genes in Bt ssp. *berliner* 1715 (Mahillon *et al.*, 1985). Nine other IS*231*A relatives have also been cloned (Figure 8.8) most of which have been found in close proximity to crystal genes on large plasmids. It therefore seems probable that IS*231* is involved in the translocation of crystal genes despite the fact that no co-transfer of a *cry* gene with IS*231* has ever been directly or indirectly recorded. All IS*231* elements (with the exception of IS*231*B and IS*231*E) are 1655 to 1657 bp long, including 20 bp IR at their ends, encode a single 477-478 amino acid putative transposase, and are flanked by 11 bp direct repeats. IS*231*B and IS*231*E are structurally inactive due to the respective disruption and deletion of one IR in each case.

IS*232* contains two overlapping orfs that span almost the entire sequence and encode 50-kDa (Orf1) and 29-kDa (Orf2) polypeptides (Menou *et al.*, 1990). Although no function has been ascribed to these proteins, Orf1 has a potential helix-turn-helix motif (Dodd & Egan, 1987) in the N-terminal region with similarity to DNA-binding proteins implicated in transposition, regulation of gene expression, or site specific recombination. As no sequence duplication has been noted at the insertion site of IS*232*A, it was suggested that this IS is not the functional element in the *cry*1A composite transposon (Menou *et al.*, 1990).

As was mentioned in the introduction to this Chapter, two copies of IS*240* (differing by only 6 bp) were found to flank the 125-kDa *cry4A* gene on the 72 MDa plasmid of Bt ssp. *israelensis* (Bourgouin *et al.*, 1988; Delécluse *et al.*, 1989). A single Orf was found to span 81% of IS*240*, encoding a protein of 28-kDa related to the functional transposases from IS*26* and IS*S1* (Delécluse *et al.*, 1989). Although a perfect inverted repeat was found in IS*240*B, no direct repeats were identified flanking the element and so no information regarding the tranpositional activity of this element is available (Delécluse *et al.*, 1989). Since this initial discovery, Rosso and Delécluse (1997) have investigated the presence of IS*240* in 69 Bt strains. Restriction digests of total DNA and PCR products obtained with a single 16-bases primer corresponding to the IS*240* inverted repeated sequence were hybridised with the IS*240*A element. The results indicate that 67% of the Bt strains tested, including all known mosquitocidal strains, possess at least one IS*240*-related element. PCR experiments indicated that IS*240* represents a family of insertion sequences with several variants (Rosso & Delécluse., 1997).

NAME	FAMILY	ORIGIN	ACCESSION NUMBER	LENGTH (bp)	IR (bp)	DR (bp)
IS*231*A	IS*4*	ssp. *berliner* 1715	X03397	1656	20	10, 11, 12
IS*231*B	IS*4*	ssp. *berliner* 1715	M16158	1646	7/20	ND
IS*231*C	IS*4*	ssp. *berliner* 1715	M16159	1655	19/20	11
IS*231*D	IS*4*	ssp. *finitimus*	X63383	1657	17/20	ND
IS*231*E	IS*4*	ssp. *finitimus*	X63384	1616	0/20	ND
IS*231*F	IS*4*	ssp. *israelensis*	X63385	1655	19/20	12
IS*231*G	IS*4*	ssp. *darmstadiensis*	M93054	1649	20	ND
IS*231*H	IS*4*	ssp. *darmstadiensis*	M93054(P)	817	ND	ND
IS*231*V	IS*4*	ssp. *israelensis*	M83546	1964	21/22	ND
IS*231*W	IS*4*	ssp. *israelensis*	M86926	1964	21/22	ND
IS*232*A	IS*21*	ssp. *berliner* 1715	M38370	2184	28/37	ND
IS*232*B	IS*21*	ssp. *kurstaki* HD73	M77344(P)	2200	28/37	ND
IS*232*C	IS*21*	ssp. *kurstaki* HD73	ND	2200	28/37	ND
IS*240*A	IS*6*	ssp. *israelensis*	J03315	865	16	ND
IS*240*B	IS*6*	ssp. *israelensis*	J03315	865	16	ND
ISBT*1*	IS*3*	ssp. *aizawai* HD229	L29100	999	15/17	ND
ISBT2	IS*3*	YBT-226	R1	187	ND	ND
Tn*4430*	Tn*4430*	ssp. *kurstaki* KTo	X13481	4149	38	5
Tn*5401*	Tn*4430*	ssp. *morrisoni* EG2158	U03554	4837	53	5

Figure 8.8 Transposable elements from *Bacillus thuringiensis*.
ND, not determined; P, partial sequence; R1, Hodgman *et al.* (1993); IR, inverted repeat; DR, direct repeat. The names ISBT*1* and ISBT2 were arbitrarily chosen and do not correspond to official denominations from the Plasmid and Transposon Reference Centre.
This table was taken and modified from Mahillon *et al.* (1994).

It was therefore not surprising that a putative transposase (*orfA*), showing 29.2% sequence similarity to that of IS*240*A was identified in the dipericidal Bt ssp. *fukuokaensis* 84-I (Dunn & Ellar, 1997). OrfA is flanked by a 17 bp imperfect inverted repeat. However, the absence of direct repeats and more importantly the presence of a stop codon which interrupts the long open reading frame, indicates that this element is unlikely to be still active. Approximately 300 bp downstream of *orfA* and orientated in the opposite direction, two overlapping open reading frames were found (*orfX* and *orfY*). Database searches revealed similarities to the IS3 family of insertion sequences. The possession of tandem, out of phase orfs is a particular characteristic of these elements. 120 bp upstream of *orfX* an *orfRec* (interrupted by a single stop codon) was identified. The site-specific recombinases to which this gene shows similarity forms part of the resolvase-invertase family which plays a role in the resolution of co-integrate molecules formed in replicative transposition.

The identification of the vestiges of these two novel insertion sequences and *orfRec* (previously found as a functional part of Class II transposons in *S. aureus*) upstream of the putative chaperone *orf1* suggested that this region of DNA has been a hot spot for

transposition-mediated recombination in the history of this organism (Dunn & Ellar, 1997). This would fit with the hypothesis presented in Chapter 6 that *orf1* may have originally been associated with *cry20A* from the same strain before it was transposed to a new genetic locus.

IS*240* like elements have also been found associated with the newly identified *cry19A* gene from Bt ssp. *jegathesan* (Rosso & Delécluse, 1997b). The similarity of *cry19A* to the *cry4A* gene of Bti (see Chapter 7) from which IS*240* was originally identified indicates a mechanism whereby duplication of an ancestral progenitor of these genes could have occurred, eventually leading to the evolution of two distinct but obviously related genes.

Smith *et al.*, (1994) identified a 999 bp element positioned downstream of *cry11A* and associated with *cry10A*. Comparison of its 262 amino acid putative transposase with other transposable elements revealed striking similarities with IS belonging to the IS3 family and especially with IS*150* from *E. coli* with which it shares 35% identity. Hodgman *et al.* (1993) also identified another IS*150* element located downstream of the cryptic *cryII* gene from YBT-226. No information is available on the *in vivo* activity and possible participation of these two elements in transfer of crystal genes.

8.5.3 Class II Mobile Elements in Bt

The first Class II transposon (Tn*4430*) has been found both on plasmid (mostly those associated with *cryIA* genes) and chromosomal DNA (Green *et al.*, 1989; Lereclus *et al.*, 1983; Mahillion *et al.*, 1988; Sanchis *et al.*, 1988)and has been shown to be active both in its original host and to participate in the transfer of plasmids pXO1 and pXO2 from *Bacillus anthracis* (Green *et al.*, 1989). In addition, the ability of Tn*4430* to transpose in *E. coli* has facilitated the functional characterisation of this element (Mahillion & Lereclus, 1988).

As for other Class II elements, Tn*4430* mobility is governed by two functional modules in a two-step reaction. In the first step, the replicative transposition of the element from its original site to the target, forms a cointegrate molecule containing two Tn*4430* copies in the same orientation at the donor-target molecule junctions. This activity requires the presence of 987 amino acid transposase TnpA acting on the 38 bp inverted repeats. During this process a 5 bp duplication (direct repeat) is generated at the target site. In the second step, the cointegrate is resolved through site-specific recombination between two resolution 'res' sites, one from each copy of Tn*4430*. In previously described transposons this second step has been found to be carried out by a 200 amino acid resolvase (TnpR). Tn*4430* uses a 284 amino acid protein (TnpI) which belongs to the other family of site-specific recombinases, that of the integrases (Mahillon & Lereclus, 1988).

The second Class II transposon Tn*5401* was isolated as a spontaneous insertion in a recombinant plasmid from Bt ssp. *morrisoni* (Baum, 1994). Like Tn*4430*, this element has

also been shown to transpose to both chromosomal and plasmid DNA and resolution of cointegrate molecules is mediated by TnpI protein (related to the integrase family).

8.5.4 Analysis of the Putative Transposase T1 Located Downstream of *am1*

As described in Section 8.3 analysis of the nucleotide sequence downstream of *am1* revealed the presence of an open reading frame encoding a putative 229 amino acid protein with a deduced molecular weight of 27272 Da. Its complete amino acid sequence is shown in Figure 8.2. The EMBL and SWISSPROT databases were searched for sequences similar to both the DNA and protein sequence of the predicted Orf (*T1*) using the BLAST program (Section 2.1.6). Similarities were found to putative transposases from Gram positive bacteria including *Lactococcus lactis* ssp. *lactis* IS*214*-II, *Enterococcus faecium* IS*1216*V and *Staphylococcus aureus* IS*257*-2 (accession numbers are shown in Figure 8.8). These transposases are all members of the IS*6* family of insertion sequences, a homologous group of about thirty elements whose archetypes are IS*26* from *Proteus vulgaris* (Mollet *et al.*, 1983) and IS*15*D from *Salmonella panama* (Trieu-Cuot & Courvalin, 1984). A subset of this group are the IS*240* insertion sequence elements in *Bacillus thuringiensis* (Delécluse *et al.*, 1989; 1990). Figure 8.9 shows the percentage similarity of T1 to the transposases of several elements in the IS*6* family and in Figure 8.10 the relationship between this Orf and other Bt transposases is shown. It is clear that of the Bt insertion sequences thus far identified, the transposase from IS*240*A displays the closest similarity to T1 (approximately 38.5% similarity). The most divergent Bt transposase to T1 is that of IS*232*A (displaying only 11.5% similarity).

The transposases from the Gram positive bacteria shown in Figure 8.9 to which T1 is related (as determined by BLAST searches for homology) comprise a family of Insertion Sequences which share common features; their length (between 800 and 865 bp) and the size of the polypeptide encoded (between 224 and 235 amino acids). The lengths of the terminal IR vary among IS*240*, ISS*1* and IS*431* (Delécluse *et al.*, 1990), however, there are eleven identical bases in all cases, as indicated by asterisks in bold text below;

```
IS240B          AAGGTTCTGGTGCAAA
ISS1              GGTTCTGTTGCAAAGTTT
IS431L        AATTCTGGTTCTGTTGCAAAGT
                  **  *** ******
```

Percentage Similarity

	T1	IS*214*-II[1]	IS*1216*V[2]	IS*257*-2[3]	IS*240*A[4]	
T1		58.1	58.3	57.0	37.6	T1
IS*214*-II	37.5		96.9	59.7	38.3	IS*214*-II
IS*1216*V	38.2	1.8		59.3	38.6	IS*1216*V
IS*257*-2	42.0	38.8	49.5		34.8	IS*257*-2
IS*240*A	59.6	58.0	58.1	61.4		IS*240*A
	T1	IS*214*-II	IS*1216*V	IS*257*-2	IS*240*A	

Percentage Divergence

Figure 8.9 Protein sequence relationships between T1 and the transposases of several elements in the IS*6* family of Insertion Sequences using a ClustalV PAM250 residue weight table.
EMBL accession numbers:- 1: Y10522; 2: L40841; 3: B60634; 4: JO3315

Percentage Similarity

	T1	IS*240*A	ISBT*1*	IS*231*A	IS*232*A	
T1		38.5	12.5	12.0	11.5	T1
IS*240*A	57.4		11.9	11.1	11.5	IS*240*A
ISBT*1*	84.5	80.2		13.7	11.5	ISBT*1*
IS*231*A	84.9	83.3	77.4		10.7	IS*231*A
IS*232*A	88.8	86.1	83.1	84.4		IS*232*A
	T1	IS*240*A	ISBT*1*	IS*231*A	IS*232*A	

Percentage Divergence

Figure 8.10 Protein sequence relationships between T1 and other Bt transposases using a Clustal PAM250 residue weight table.
EMBL accession numbers as shown in Figure 8.8

Although the IR of these elements are highly conserved, no directly repeated sequences suggestive of target duplication have been noted in IS*240*. The absence of direct repeats could be either due to rearrangements of the regions surrounding one or both IS*240* elements, or to a transposition event which does not generate a target duplication. The latter hypothesis seems unlikely since the related ISS*1* and IS*431*L produce 8 and 4 bp duplication at the insertion site respectively.

A multiple sequence alignment between T1 and the related putative transposases from Gram positive IS with greater than 40% similarity is shown in Figure 8.11. It is clear that the sequences are highly conserved thoughout the length of the protein. Figure 8.12 shows the results of a similar alignment of T1 with the putative transposase of IS*240A*. It is interesting to note that in both alignments, the invariant motif D, D (35), E is conserved. This motif has been found in the integrase proteins (IN) of Rous sarcoma virus and human immunodeficiency virus. Its *in vitro* mutation results in only 1 to 8% of wild type activity (Kulkosky *et al.*, 1992). It is thought that these acidic residues are involved in binding the metal ions Mg^{2+} or Mn^{2+} which are essential in both the processing and joining reactions of IN (Kulkosky *et al.*, 1992).

As has already been discussed above, putative transposable elements have recently been identified in the closely related strain Bt ssp. *fukuokaensis* 84-I (Dunn & Ellar, 1997) in association with the *orf1* gene. One of these elements also showed similarity to other Gram positive insertion sequences (IS) including 29.2% similarity to IS*240*. The deduced amino acid sequence of the putative transposase of this insertion sequence is referred to as Orfα and encodes a 227 amino acid protein incorporating a stop codon (TGA) where amino acid 71 should be (as determined by ClustalV multiple sequence alignments). This probably indicates that this transposable element is a vestige of a once functional IS which is now incapable of independent transposition. Further evidence to support this comes from the lack of direct repeat duplications which should be present adjacent to the imperfect 17 bp terminal inverted repeats

A ClustalV alignment of the amino acid sequences of Orfα from 84-I and T1 from 17A shown in Figure 8.13 shows the high degree of sequence conservation (73.6% similarity) between these two elements. It is therefore possible that these two putative IS, could be considered to be part of a new class of Bt insertion sequences.

```
T1          M--RYFKGKQFKKDIILVAVGYYCRFSLSYRDVSEILRERGISVHPTTIMRWVHEYGNQM
IS214/II    MTMNHFKGKQFQQDVIIVAVGYYLRYNLSYREVQEILYDRGINVSHTTIYRWVQEYGKLL
IS1216V     MTMNHFKGKQFQQDVIIVAVGYYLRYNLSYREVQEILYDRDINVSHTTIYRWVQEYGKLL
IS257/2     M--NYFRYKQFNKDVITVAVGYYLRYALSYREISEILYERGINHSSTVIYRWVQEYAPVL
            *    *  ***  *.* ****** *  ****.. *** .* ..*  .*. *** **

T1          YQIWKKKNKSAQLSWHLDETYIKVKGEWCYLYRAIDKDGHTLDIQLRKKRDHQAAYAFMK
IS214/II    YQIWKKKNKKSFYSWKMDETYIKIKGKWHYLYRAIDADGLTLDIWLRKKRDTQAAYAFLK
IS1216V     YQIWKKKNKKSFYSWKMDETYIKIKGKWHYLYRAIDADGLTLDIWLRKKRDTQAAYAFLK
IS257/2     YQIWKKKHKKAYYKWRVDETYIKIKGQWCYLYRAIDADGHTLDIWLRKKRDHQSAYAFIK
            ******* * .   *  ******.** * ******* ** **** ****** *.**** *

T1          RLVKTFGEPTDLTTDKAPALLCAFNKL-REQGFYKHTTHCTIKHLNNLIEQDHRHIKRRF
IS214/II    RLVKQFDEPKVVVTDKAPSIASAFKKL-KEYGFYQGTEHRTIKYLNNLIEQDHRPVKR--
IS1216V     RLVKQFDEPKVVVTDKAPSIASAFKKL-KEYGFYQGTEHRTIKYLNNLIEQDHRPVKR--
IS257/2     RLIKQFGKPQMIITDQAPSTKVAMAKIIKVFKL-KPNCHCTSKYLNNFIEQDHRPIKV--
            **.* *  *    ** **.   *  *           * * * *** ******  .*

T1          SRSAGFQSLRHASRTLKGIETIHAIYKRKRSLQQPNFAFSTYNELQQFLTIA
IS214/II    -RNKFYRSLRTASPTIKGMEAIRGLYKKTRK-EGTLFGFSVCTEIK-LLGIPD
IS1216V     -RNKFYRSLRTASPTIKGMEAIRGLYKKTRK-EGTLFGFSVCTEIKVLLGIPD
IS257/2     -RKTRYQSVNTAKNTLKGIECIYGLYKKT----VGLFRSTDFHHATKL--VIC
             *    *   * * ** * *   **        *  .          .
```

Figure 8.11 ClustalV protein sequence alignment of T1 with putative Gram positive transposases.
See Figure 8.9 for accession numbers. Identical residues are marked with an asterisk and . denotes similar residues. The invariant motif D, D, E is shown in bold.

```
T1          M---RYFKGKQFKKDIILVAVGYYCRFSLSYRDVSEILRERGISVHPTTIMRWVHEYGNQ
IS240-A     MEKENIFKWKHYQADMILWTVRWYLRYNLSFRDLVEMMEERGLSLSHTTIMRWVHQYGPE
            *     ** *    * ** .* * * .** **  *   *** *   ******** **

T1          MYQIWKKKNKSAQLSWHLDETYIKVKGEWCYLYRAIDKDGHTLDIQLRKKRDHQAAYAFM
IS240-A     LNERIRKHLKRTNDSWRMDETYIKIKGENMYLYRAVDSEGNTLDFYLSKKRDEKAAKCFL
                * *  .  **  ******.***  *****.* .* ***  * ****  **  *

T1          KRLVKTF--GEPTDLTTDKAPALLCAFNKLREQGFYKHTTHCTIKH-LNNLIEQDHRHIK
IS240-A     KKALASFHVTKPRVITVDGNKAYPVAIRELKNEKSISYGMPLRVKKYLNNMIEQDHRFIK
            *    .*    *   * *   *   *   *                .*  *** ****** **

T1          RRFSRSAGFQSLRHASRTLKGIETIHAIYKRKRSLQQPNFAFSTYNELQQFLTIA
IS240-A     KRIRNMLGLKSMQTAVKMIAGIEAMHMVKKGQLKLRAQSAQNQNRCIHQLFGLTA
             *    *  *   *     ***. * . *    *   .         * *   *
```

Figure 8.12 ClustalV protein sequence alignment of T1 and the putative transposase from IS*240A*.
See Figure 8.8 for accession number. Identical residues are marked with an asterisk and . denotes similar residues. The invariant motif D, D, E is shown in bold.

```
T1      MRYFKGKQFKKDIILVAVGYYCRFSLSYRDVSEILRERGISVHPTTIMRWVHEYGNQMYQ
Orfα    MGYFKGIQFKKDIILVAVGYYCRFSLSYRDVFEILKENVVSVHPTTIMRWVHEDGQLIYQ
        * **** ************************ *** *  .************* *   **

T1      IWKKKNKSAQLSWHLDETYIKVKGEWCYLYRAIDKDGHTLDIQLRKKRDHQAAYAFMKRL
Orfα    IWKKK--IAHQIZHLDETYIKVKGKWRYLYHAINGEGHTLDIQLRRTRNHEAAYACMKRL
        *****  .*  . **********  * *** **. .*********  *.* **** ****

T1      VKTFGEPTDLTTDKAPALLCAFNKLREQGFYKHTTHCTIKHLNNLIEQDHRHIKRRFSRS
Orfα    HKHFGEPTVHTTDKAPPLLCAFKKLRNSSFYTHTIHCTVKHLNNFIERDHRHVKRCFVKS
         * *****  ****** ***** ***    ** ** ***.***** ** ****.** *  *

T1      AGFQSLRHASRTLKGIETIHAIYKRKRSLQQPNFAFSTYNELQQFLTIA
Orfα    TGFQTLRHTSLNLKGIETVHVLYKRKRSLQQPNFVFSIYNELQQLLAIA
        .***.***.*  ******.*  ************ ** ****** *.**
```

Figure 8.13 ClustalV protein sequence alignment of T1 and Orfα from Bt ssp. *fukuokaensis* 17A and 84-I respectively.
Identical residues are marked with an asterisk and . denotes similar residues. The invariant motif D, D, E is shown in bold.

A potential ribosome binding site GAGGGG was identified 8 bp upstream of the T1 ATG initiation codon (Figure 8.14). The sequence was also investigated for the presence of potential stem-loop structures. A region of perfect dyad symmetry had been found just before the end of the coding sequence in Orfα. Study of the same region in T1 revealed a very similar sequence containing a single mismatch (Figure 8.14). An imperfect potential stem loop structure had also been found immediately downstream of Orfα. Sequence downstream of *T1* showed little if any similarity to that of Orfα but a potential stem loop structure was identified (Figure 8.14) with perfect symmetry for at least 6 bp just after the deduced TAA stop codon.

```
AACTGCCCCGTTAGTTCAATAAGAAAAAAATTAGGTTCTGTTGCAAAGATGTCTAAGCAA 60
                                          -------
                                            -35
                                         M  R  Y  F  K  G  K
AGTTAAAATTTAAAGAAAATGACTCGAGGGGCATGATACATGAGATATTTTAAAGGAAAA 120
   ------
    -10
Q  F  K  K  D  I  I  L  V  A  V  G  Y  Y  C  R  F  S  L  S
CAATTCAAGAAAGATATTATTTTGGTAGCCGTTGGCTATTACTGTCGTTTTTCTTTAAGT 180

Y  R  D  V  S  E  I  L  R  E  R  G  I  S  V  H  P  T  T  I
TATCGTGATGTATCTGAAATTTTGAGGGAACGTGGCATCTCTGTTCATCCGACAACGATT 240

M  R  W  V  H  E  Y  G  N  Q  M  Y  Q  I  W  K  K  K  N  K
ATGCGTTGGGTACATGAATATGGAAATCAGATGTATCAAATTTGGAAAAAGAAAAACAAA 300

S  A  Q  L  S  W  H  L  D  E  T  Y  I  K  V  K  G  E  W  C
TCCGCACAGCTATCTTGGCACCTGGATGAAACCTATATCAAAGTGAAAGGCGAATGGTGT 360

Y  L  Y  R  A  I  D  K  D  G  H  T  L  D  I  Q  L  R  K  K
TATCTCTATCGAGCAATCGATAAAGATGGACATACATTGGATATTCAACTTCGCAAAAAA 420

R  D  H  Q  A  A  Y  A  F  M  K  L  L  V  K  T  F  G  E  P
CGAGATCATCAAGCTGCATATGCCTTTATGAAAAGACTAGTGAAAACATTTGGAGAACCA 480

T  D  L  T  T  D  K  A  P  A  L  L  C  A  F  N  K  L  R  E
ACGGATCTCACAACAGACAAAGCGCCCGCGCTACTTTGTGCGTTCAATAAATTGAGAGAA 540

Q  G  F  Y  K  H  T  T  H  C  T  I  K  H  L  N  N  L  I  E
CAAGGCTTTTATAAGCATACCACTCATTGTACAATCAAGCATTTGAATAATCTCATTGAA 600

Q  D  H  R  H  I  K  R  R  F  S  R  S  A  G  F  Q  S  L  R
CAGGATCATAGACATATCAAGCGACGTTTTTCCAGATCCGCAGGATTTCAAAGTCTTCGC 660

H  A  S  R  T  L  K  G  I  E  T  I  H  A  I  Y  K  R  K  R
CATGCTTCACGCACCTTGAAAGGAATCGAAACAATTCATGCTATATATAAACGAAAACGA 720
                                                    =>=>=>

S  L  Q  Q  P  N  F  A  F  S  T  Y  N  E  L  Q  Q  F  L  T
AGTTTGCAACAACCAAACTTCGCTTTTTCGACGTACAATGAATTACAACAATTTCTAACG 780
=>=>=>        <=<=<=<=<=<=<=<=

I  A  *
ATCGCATAAAATCTATCTGCAATCATACCAACTTTTTATCTGTAAAGAAACTTTGCAACA 840
      ->->->->->->->          <-<-<-<-<-<-<-

GAACCTCCAGTAGTTATGGGGTACGGAAAGAATGGCATATTAAA
```

Figure 8.14 Nucleotide and amino acid sequence of the *T1* open reading frame.
The 5' inverted repeat sequence is indicated in bold. Regions of dyad symmetry located both upstream and downstream of the TAA stop codon which may act as termination signals are indicated by opposing arrows. The potential ribosome binding site proximal to the first ATG initiation codon of the *T1* reading frame is marked in bold and underlined. The dashed lines indicate putative upstream -35 and -10 control sequences (as described in Chapter 5).

Using the nucleotide interpretation program (Staden, 1987) a search was conducted for inverted repeats characteristic of transposable elements. A perfect 16 bp IR sequence was found from base 33 to 48 and from base 830 to 846 (Figure 8.14). A very similar 17 bp IR sequence had been found adjacent to Orfα but this sequence contained a single mismatch at position 3 in the sequence. The inverted repeat sequence of T1 also aligns with the reported IR of other Gram positive IS6-like insertion sequences (Figure 8.15).

```
IS240B          AAGGTTCTGGTGCAAA
ISS1              GGTTCTGTTGCAAAGTTT
IS431L        AATTCTGGTTCTGTTGCAAAGT
ISOrfα            GGATCTGTTGCAAAGTT
IST1              AGGTTCTGTTGCAAAG
                  **  *** ******
```

Figure 8.15 Inverted repeat sequences of IS6-like insertion sequences.
Conserved residues are marked with an asterisk in bold text.

Consensus promoter sequences in Bt responsible for directing transcription during sporulation have been discussed in Chapter 5 (Brown & Whiteley, 1988; Baum & Malvar, 1995) and were identified upstream of *am1*. Sequence directly upstream of *T1* was aligned with that of *Orfα* and other IS6 transposases in an attempt to locate *T1* potential promoter sequences. A potential -35 Pribnow box was easily identified at position 41 (Figure 8.14) by its identity to the -35 box sequence of other IS6 transposases (Figure 8.16). Alignment of the -35 regions then facilitated identification of an AT rich region (Figure 8.14) that aligned with similar regions believed to be -10 Pribnow boxes in other IS6 elements (Figure 8.16). The -35 sequence lies within the 5' inverted repeat flanking *T1* as it does for *orfα*. Figure 8.16 shows the alignment of these potential control sequences with those of ISOrfα IS431 (a similar element to IS257/2; Barberis-Maino *et al.*, 1987) and ISS1 (Polzin & Shimizu-Kadota, 1987).

```
                -35                          -10

IS431L   TTGCAAAGTTGAATTGAATTTATAGTATAATTTTAACAAAAAGGA
         -------                     ------

ISS1     TTGCAAAGTTTTCGGATAAGTCTATTTTAGTGTAAAATGAATAAAA
         -------                     ------

ISOrfα   TTGCAAAGTTCTTTTCACCTGATTACAATTAGAAAAAAGATACAAG
         -------                     ------

IST1     TTGCAAAGATGTCTAAGCAAAGTTAAAATTTAAAGAAAATGACTC
         --------                    ------
```

Figure 8.16. Alignment of potential promoter sequences lying upstream of the putative *T1* initial ATG codon and those of IS*431*, IS*S1* (Barberis *et al.*, 1987; Polzin & Shimizu-Kadota, 1987) and IS*Orfα* (Dunn & Ellar, 1997).

It is interesting to note the long distance between the ATG start codon and the -10 box in all of the above potential promoter structures. IS elements require a strict control on their mobility to prevent high levels of transposition which would be destructive to their host (Iida *et al.*, 1983) and this structure may represent a very weak promoter assisting transcriptional regulation of the transposition. Possible stem-loop conformations containing the potential RBS have been noted in the region upstream of several IS elements (Mahillon *et al.*, 1985; Timerman & Tu, 1985; Klaer *et al.*, 1981; Kröger & Hoborn, 1982; Halling *et al.*, 1982 & Dalrymple *et al.*, 1984) including the putative IS*231* transposase gene from Bt ssp. *berliner* 1715 (Mahillon *et al.*, 1995) and are believed to contribute to translational control. This stem-loop conformation will only be formed on mRNA transcribed from a promoter upstream of the IS and could thereby entrap the RBS so as to prevent fortuitous translation of the putative transposase. Such a structure could also function as a transcription terminator preventing read-through transcription into the IS element. The transcript from the *T1* putative promoter would not yield this structure and study of the region upstream of *T1* did not reveal any convincing stem-loop structures that could act in this way to prevent unwanted transcription of the putative transposase gene. The long distance between the ATG start codon and the -10 box may therefore be needed to assist transcriptional regulation as described above.

8.5.5 Analysis of the Putative Transposase T2 Located Upstream of *am1*

As introduced in the previous section, a second putative transposase T2 was identified immediately upstream of *am1*. T2 appears to be encoded by two overlapping open reading frames *orfX* and *orfY* (shown between nucleotides 7665 to 7356 and 8046 to 7548 respectively in Figure 8.3). As previously discussed, alignment of *orfX* and *orfY* with IS*240A* (the sequence found to have the highest degree of similarity to *orfX* and *orfY*) indicated that a single frame shift mutation (marked in Figure 8.3) is responsible for the disruption of the open reading frame encoding T2. Alignment of the two regions of *orfX* and *orfY* corresponding to IS*240A*, with IS*240A* and other Bt transposases using a Clustal IV alignment generated the similarity values shown in Figure 8.17. It is obvious from this data that T2 is indeed highly related to IS*240A*. This represents yet another identification of an IS*240* type sequence within a diptericidal Bt strain (Rosso & Delécluse, 1997) and further supports the theory discussed in previous Chapters that these mobile elements have been involved in the dispersal of *cry4*-type genes. An alignment of T2 with IS*240A* is shown in Figure 8.18.

Despite the fact that the putative ATG start codon of T2 aligns with the reported start of IS*240A*, a search of the region immediately upstream of T2 did not reveal any very obvious promoter sequences. The first potential ribosome binding site GGAAAA was identified 12 nucleotides upstream of the T2 start codon and although an AT rich region reminiscent of a -10 box does exist between nucleotides 8089 and 8093, no obvious region corresponding to a -35 box could be found. Nor is there any evidence that this putative insertion sequence is flanked by either inverted or direct repeat sequences. The lack of an obvious promoter region and flanking repeat sequences along with the presence of the frameshift mutation indicates that this sequence may have undergone considerable rearrangement since its initial insertion into this site and is almost certainly no longer functional.

Percentage Similarity

	T2	IS*240*A	ISBT*1*	IS*231*A	IS*232*A	
T2		80.5	13.4	12.6	11.3	T2
IS*240*A	17.8		11.9	11.1	11.5	IS*240*A
ISBT*1*	79.8	77.7		13.7	11.5	ISBT*1*
IS*231*A	81.5	81.9	77.4		10.7	IS*231*A
IS*232*A	83.8	85.6	86.0	85.9		IS*232*A
	T2	IS*240*A	ISBT*1*	IS*231*A	IS*232*A	

Percentage Divergence

Figure 8.17 Protein sequence relationships between T2 and other Bt transposases using a Clustal PAM250 residue weight table.
ISBT*1*, IS*231*A, IS*232*A and IS*240*A EMBL accession numbers are included in Figure 8.8.

```
T2        MEKENIFKWKHYQPDIILLTVRWYLRYSLSFGDLVEMTEERGLSMSYTTIMRWVHQYGPE
IS240A    MEKENIFKWKHYQADMILWTVRWYLRYNLSFRDLVEMMEERGLSLSHTTIMRWVHQYGPE
          ************** * ** ********.*** ***** ****** * **************

T2        LNERIRKHLKPTNDSWRVDETYLKIKGENMYLYRAVDSEGNTLDFYLSSKRDAKAAKRFL
IS240A    LNERIRKHLKRTNDSWRVEETYIKIKGENMYLYRAVDSEGNTLDFYLSKKRDEKAAKCFL
          ********** ********.*** ************************ *** **** **
                                ⇓
T2        KKALASCHVTKPRIFTVDZM----VMIQKLKEEKRLPHDTPLRVKKYLNTIIEQDHRFIIK
IS240A    KKALASFHVTKPRVITVDGNKAYPVAIRELKNEKSISYGMPLRVKKYLNNMIEQDHRFIIK
          ****** ******. ***       * *  ** **       *********  **********

T2  F     KRIRNMLGLKSLRTATKMIAGIEVVHMIKKGQLKLRVQSLKNQNRCIHQLFRLPA
IS240A    KRIRNMLGLKSMQTAVKMIAGIEAMHMVKKGQLKLRAQSAQNQNRCIHQLFGLTA
          ***********  ** *******  **.******** **  ********** * *
```

Figure 8.18 ClustalV protein sequence alignment of T2 with IS*240A* from Bti.
Identical residues are marked with an asterisk and . denotes similar residues. The invariant motif D, D, E is shown in bold. The frame shift mutation between *orfX* and *orfY* is denoted by the letter Z and marked with an arrow.

8.5.6 Analysis of the Putative Integrase/Recombinase Downstream of *am1*

Analysis of the sequence furthest downstream of *am1* revealed the presence of an open reading frame (*OrfD*) between nucleotides 4950 (at the end of clone pAM10) and 4613. Database searches of the deduced protein sequence of this Orf revealed some similarity to a number of integrase/recombinase proteins from a variety of different organisms. The highest levels of similarity being to XERC from *Haemophilus influenzae* and XERD from *E. coli*.

Such proteins participate in site-specific recombination by catalysing the cutting and rejoining of the recombining DNA molecules. Site specific recombinases have also been shown to function in plasmid partitioning (Swinfield *et al.*, 1991; Pujol *et al.*, 1994) where they enhance the segragational stability of plasmids by reducing the number of plasmid multimers resulting from homologous recombination. The ability to maintain the plasmid population in a monomeric state would be particularly important for large low copy number plasmids such as are found in Bt.

Dunn & Ellar (1997) had identified a putative recombinase upstream of *orf1* with over 60% similarity to *sin* (a potential recombinase-encoding gene from *S. aureus*) and *bin*L (a resolvase gene also from *S. aureus*). These two genes were found to reside on the Class II transposons Tn*4002* and Tn*552* respectively and form part of the resolvase and DNA invertase family of site-specific recombinases (Stark *et al.*, 1992).

Multiple sequence alignments of *orfD* with XERC and XERD did not reveal such high levels of identity but the alignment in Figure 8.19 shows some similarity (25.0% to XERC and 26.8% to XERD) particularly over the central region. The alignment also shows that *orfD* is orientated in the opposite direction to *am1* and is N-terminally truncated by the end of clone pAM10.

An alignment was also carried out with the highest scoring Bt integrase/recombinase, Tnp1 from the Class II transposon Tn*4430* (Figure 8.20) which was found to have 17% similarity to OrfD. Again, the region of highest similarity was found to the central region of TnpI (corresponding to amino acids 115 to 190 in Tnp1).

```
OrfD  A----ARADDVANMIFNEKD--------------DVAFLQ----------------FL
XERC  M---LTALNRYWDYLRIERQMSPHTITNYQHDLDATIKILAQQDIHSWTQVTPSVVRFIL
XERD  MKQDLARIEQFLDALWLEKNLAENTLNAYRRDLSMMVEWLHHRGL-TLATAQSDDLQALL
           **     *    **                  *  *                   *

OrfD  GNDYEEMLKDIS--------PRKFSFFQRDKE--------------------------
XERC  AESKKQGLKEKSLALRLSALRRFLSFLVQQGELKVNPATGISAPKQGRHLPKNMDGEQVQ
XERD  AERLEGGYKATSSARLLSAVRRLFQYLYREKFREDDPSAHLASPKLPQRLPKDLSEAQVE
          *  **  *          * ***  * **

OrfD  --------------RDIAIISLILGTGLPVSEVASLTISSINFRQGTVKVTRKGNK----
XERC  QLLAND-SKEPIDIRDRAILELMYSSGLRLSELQGLDLNSINTRVREVRVIGKGNKERVV
XERD  RLLQAPLIDQPLELRDKAMLEVLYATGLRVSELVGLTMSDISLRQGVVRVIGKGNKERLV
                    ** **  *   *** ***   ** **** *** * *  ****

OrfD  ------------------------------------------------RYSI-------
XERC  PFGRYASHAIQEWL---KVRALFNPKDEALFVSQLGNRISHRAIQKRLETWGIRQGLNSH
XERD  PLGEEAVYWLETYLEHGRPWLLNGVSIDVLFPSQRAQQMTRQTFWHRIKHYAVLAGIDSE
                                                      * *

OrfD  --------------------------LATRSCLDDVQEYIRVAS
XERC  -LNPHKLRHSFATHMLEASSDLRAVQELLGHSNLSTTQIYTHLNFQHLAEVYDQAHPRAKRKK
XERD  KLSPHVLRHAFATHLLNHGADLRVVQMLLGHSDLSTTQIYTHVATERLRQLHQQHHPRA
                                  *   * *   * *  **
```

Figure 8.19 ClustalV protein sequence alignment of OrfD with the integrase/recombinase proteins XERC of *Haemophilus influenzae* (Swiss-Prot P44818) and XERD of *E. coli* (Swiss-Prot P21891).
Conserved and semi-conserved residues are shown in bold and marked with an asterisk. Residues that are conserved between OrfD and one other sequence are marked with an asterisk in plain text.

```
OrfD  --------------------------AARADDVANMIFNEK-------DDVAFLQFLGN
Tnp1  MDVAKQFSSYLKQENKTENTVQGYTSGIRQYIKWFEGSYDRKLTKLYRQNILEYISYLKN
                                          * *        *         * *

OrfD  -----------------DYEEMLK-----------------------------DISPRKF
Tnp1  VKMLNAKSINHKISSLAKFNEFLIQKGSQQDQVILKTDMIKVQTVYASPTQIVELDVKKF
                         * *                                *     **

OrfD  --SFFQRDKERDIAIISLILGTGLPVSEVASLTISSINFRQGTVKVTR-KGNKRYSILAT
Tnp1  LQSVLEDNNKRNYAIATLLAYTGVRISEALSIKMNDFNLQTGECIIRSGKGGKQRIVLLN
        *    *  ** ** **   **  ***  *   *  *   *   *   ** *   **

OrfD  RSCLDDVQEYI----------------------------RVAS
Tnp1  SKVLSAIKDYLIDRKTYSTAHESPYLFISKKREKLDRTVVNRIFKSYSNVITPHQLRHFF
         *  * **                                **

OrfD
Tnp1  CTNAIEKGFSIHEVANQAGHSNIHTTLLYTNPNQLQLKNKMELL
```

Figure 8.20 ClustalV protein sequence alignment of OrfD with the Bt integrase/recombinase protein Tnp1 from the Class II transposon Tn4430.
Conserved or semi-conserved residues are shown in bold and marked with an asterisk.

8.5.7 Conclusions

In this Chapter, DNA surrounding a putative δ-endotoxin gene from Bt *fukuokaensis* 17A was analysed. Directly downstream of *am1*, and transcribed in the opposite direction, an insertion sequence (*T1*) was identified with similarity to the IS*6* family of insertion sequences. The results of a database search revealed significant similarity (up to 58%) to IS*6* insertion sequences from Gram positive bacteria. Further alignment with the recently identified Orfα from Bt ssp. *fukuokaensis* 84-I also believed to be an IS*6* insertion sequence showed over 70% similarity; the next closest Bt relative being the transposase from IS*240A* part of a subset of IS*6*. At 229 amino acids the size of the putative IS*T1* is within the size range of the IS*6* family.

T1 was also found to be flanked by a 16 bp perfect inverted repeat in which all 11 of the amino acids conserved in IS*6* IRs were identified. No direct repeat duplications were identified adjacent to the IRs, this is also the case for IS*240A* and Orfα. Despite the lack of direct repeats, which may have been lost by rearrangement of one or both of the flanking regions, *T1* exists as a complete uninterrupted open reading frame that may therefore still be capable of transposition.

A second putative transposase (*T2*) identified immediately upstream of *am1* shows much greater similarity to IS*240A* (80.5%). However, the disruption of *T2* by a single frame shift mutation and the lack of both an obvious promoter region and flanking repeat sequences probably indicates that this transposable element is now incapable of independent transposition.

The identification of the incomplete open reading frame *OrfD* with similarity to integrase/recombinase proteins involved in Class II transposition further downstream of *am1* raises the possibility that *T1* and *T2* form part of a novel Class II transposon. The presence of such a transposon would provide a mechanism to explain the diversity of Cry4 type sequences in Bt.

Transposons mediate a variety of genetic arrangements including duplications, deletions, inversions and translocations of neighbouring sequences. It is therefore possible that this genetic locus has been subject to genetic rearrangements which may explain the interruption in the insertion sequence *T2* and its lack of repeats and promoter sequences. It may also provide a mechanism for the rearrangement of two Orf to form a single *cry4*-type sequence as discussed in previous Chapters. Studies on Bt transposable elements have shown that certain elements tend to insert into or very close to other mobile elements including themselves (Ryan *et al.*, 1993; Menou *et al.*, 1990). It would appear that such a situation may have occurred in this instance as several genes found in very different mobile elements are located in close proximity.

Further work should be carried out to obtain a clone containing the sequence further downstream of the *am1* gene. DNA sequence analysis of this region would confirm the

possibility that *orfD* is part of a novel Class II transposon. As such it will contain regions required for site-specific resolution and be flanked by terminal inverted repeats.

Bacillus thuringiensis has proved to be a rich source of transposable elements (reviewed by Mahillon *et al.*, 1994). Their association with δ-endotoxin genes has been demonstrated in a variety of different strains for example, ssp. *aizawai* HD229 (Smith *et al.*, 1994) and ssp. *kurstaki* HD73 (Menou *et al.*, 1990). Their involvement in toxin mobilisation is speculative but the identification of another IS*240* like insertion sequence adjacent to a Cry4 type sequence must strengthen the view that transposition does play a direct role in dissemination of *cry* sequences among Bt populations.

8.6 Small Acid-Soluble Spore Proteins (SASP)

8.6.1 Background Information

This unique class of low molecular weight (6,000 to 12,000 Da) proteins which are soluble in acetic acid were first identified and characterised by Setlow (Setlow, 1975a) from the spores of *Bacillus megaterium.* Since then, one or more major SASP has been purified and/or partially characterised from spores of several different *Bacillus* species including *B. cereus* (Yuan *et al.*, 1981), *B. subtilis* (Connors *et al.*, 1986), *B. thuringiensis* (Leighton, personal communication), *Clostridium* species (Setlow & Waites, 1976) and other sporulating bacteria as diverse as *Thermoactinomyces thalophilus* (Sun & Setlow, 1987) and *Sporosarcina* (Magil *et al.*, 1990).

In all these organisms except the lostridia two distinct types of major SASP have been identified. These two types are named for the SASP of these types in *B. subtilis*, an α/β-type and a γ-type (Johnson & Tipper, 1981). In spores of any given organism there are two major SASP of the α/β-type, as well as many minor α/β-type SASP, each coded for by a unique gene (Fliss *et al.*, 1985; Johnson & Tipper, 1981; Loshon *et al.*, 1986; Setlow, 1985; Yuan *et al.*, 1981). The α/β-type SASP have molecular weights of 5-7000 Da. They have extremely similar amino acid sequences, are closely related immunologically and have a significant percentage of hydrophobic amino acids (Connors *et al.*, 1986; Fliss *et al.*, 1985, 1986; Loshon *et al.*, 1986). The high degree of homology between these proteins suggests that they may have arisen from a common genetic ancestor by a process of gene duplication. In contrast, all organisms so far examined contain only a single γ-type SASP (and a single γ-type SASP gene) (Hackett *et al.*, 1986; Hackett & Setlow, 1987; Sun & Setlow, 1987) with a molecular weight of 8-11,000 Da (Hackett & Setlow, 1987; Setlow & Ozols, 1980; Sun & Setlow, 1987). γ-type SASP have a very different amino acid sequence from α/β-type SASP (Setlow, 1985; Sun & Setlow, 1987), do not react immunologically with antisera to α/β-type SASP (Connors *et al.*, 1986; Hackett & Setlow, 1987) and are extremely low in

large hydrophobic amino acids (Sun & Setlow, 1987). Despite the many differences, γ-type SASP has at least some sequence homology to the α/β-type (Setlow *et al.*, 1980) indicating that they are all related species.

In vivo the SASP are synthesised in the forespore (Dignam & Setlow, 1980) during a limited time in sporulation (Dignam & Setlow, 1980). During this period they comprise at least 5% of the total protein synthesised. SASP synthesis therefore differs from the synthesis in *B. subtilis* of the spore coat protein (another spore-specific protein), whose synthesis is spread out over a large period of time in sporulation (Munoz *et al.*, 1978) and takes place in the mother cell (Munoz *et al.*, 1978; Nakayama *et al.*, 1978; Cheng & Aronson, 1977).

Detailed studies carried out on *B. megaterium* have shown that up to 20% of the total dormant spore protein is degraded to free amino acids in the first minutes of spore germination of which SASP contribute more than 90% (Setlow, 1975b; Setlow & Waites, 1976). Identification of a role for such rapid and large scale degradation of SASP protein can be found on examining the requirements of spore germination.

Spores of *Bacillus* species are one of the most quiescent stages of growth known, carrying out no detectable energy metabolism or macromolecular biosynthesis (Church & Halvorson, 1957; Sakakibara *et al.*, 1965; Setlow & Kornberg, 1970). However, these dormant spores can be converted into a growing cell via the process of spore germination, and within minutes after germination has been initiated, energy metabolism and RNA and protein synthesis begin (Setlow & Kornberg, 1970; Setlow & Primus, 1975). The high energy phosphate needed in the first 15 minutes of germination can be generated wholly from endogenous energy and phosphate reserves (Setlow & Primus, 1975b). Nucleotide biosynthesis is absent during these first minutes due to the lack of biosynthetic enzymes but, RNA synthesis is supported completely in the first 15 minutes by endogenous nucleotides, most of which are produced by the breakdown of pre-existing RNA. *Bacillus* spores are known to contain a complete apparatus for protein synthesis which begins within the first few minutes of germination despite the absence of most free amino acids in dormant spores and the absence of several enzymes of amino acid biosynthesis. The amino acids generated by the rapid proteolysis of SASP described above therefore provide essential building blocks for protein synthesis early in germination (Setlow, 1975a).

Such bursts of protein degradation have also been associated with the breaking of dormancy or the activation of metabolically quiescent stages of growth in other organisms. The rapid proteolysis that accompanies the fertilisation of sea urchin eggs (Monroy *et al.*, 1965), seed germination (Dure *et al.*, 1973) and fungal spore germination (Leighton *et al.*, 1970) are all accompanied by such bursts of protein degradation and have been associated with generation of amino acids for protein synthesis by the developing organism.

Not surprisingly then, given this role as a source of amino acids during spore germination, SASP are extremely sensitive to proteolysis. This protease sensitivity, especially to non-specific proteases is probably due to a combination of absence of disulphide bonds, low molecular weight, and a high hydrophobic character. A protein with these properties would be expected to have a rather flexible conformation with many peptide bonds accessible to proteolytic attack. *In vivo* as well as *in vitro*, SASP degradation is initiated by a highly sequence specific endoprotease termed GPR found only in dormant spores and developing forespores (Postemsky *et al.*, 1978; Setlow & Kornberg, 1969; Singh *et al.*, 1977).

GPR cleaves within a peptide sequence that is similar, although not identical in both α/β-type and γ-type SASP. All major SASP contain either one (α/β-type SASP) or two (γ-type SASP) of these protease cleavage sites (Setlow *et al.*, 1980; Yuan *et al.*, 1981). Consequently, spore protease cleavage generates large (20-35 residue) oligopeptides, which are rapidly degraded by peptidases *in vitro*, and presumably *in vivo*, to free amino acids (Setlow, 1985).

GPR is synthesised as an inactive precursor (P_{46}) in the developing forespore slightly before the synthesis of its SASP substrates (Sanchez-Salez & Setlow, 1993; Setlow, 1988; Sussman & Setlow, 1991) and processed to an active enzyme (P_{41}) about 2 hours later in sporulation by a single proteolytic cleavage (Sanchez-Salez & Setlow, 1993). This autoprocessing is stimulated by the fall in pH, accumulation of DPA and forespore dehydration that accompanies forespore development (Maghill *et al.*, 1994; Sanchez-Salez & Setlow, 1993; Illades-Aguiar & Setlow, 1994).

A major issue concerning the SASP is their function in the dormant and germinating spore. It seemed unlikely that if the generation of amino acids were the sole function of these proteins that there would be so many of them or that there would be such a high concentration in the spore core (about 30 mg/ml total (Setlow, 1975c). Consequently, it was speculated that these multiple protein species serve some key function or functions in dormant spores, which can be done away with upon spore germination.

SASP degradation as a source of amino acids for protein synthesis explains the high conservation of sequence around the initial protease cleavage site (see Figure 8.21 and 8.22). However, it does not explain why if this were the sole function of SASP there is a second more C-terminal region of conserved sequence (Figure 8.21). The findings that SASP are associated with spore DNA *in vivo* and appear at the same time that spores become resistant to UV light led to the suggestion that this secondary function might be a direct role in spore UV resistance (Setlow, 1978; Setlow, 1985).

That this early suggestion was correct, at least for α/β-type SASP, has been demonstrated using strains with deletion mutants in one or more of the genes coding for the three major *B. subtilis* SASP (Mason & Setlow, 1986; 1987). Spores lacking α/β SASP had

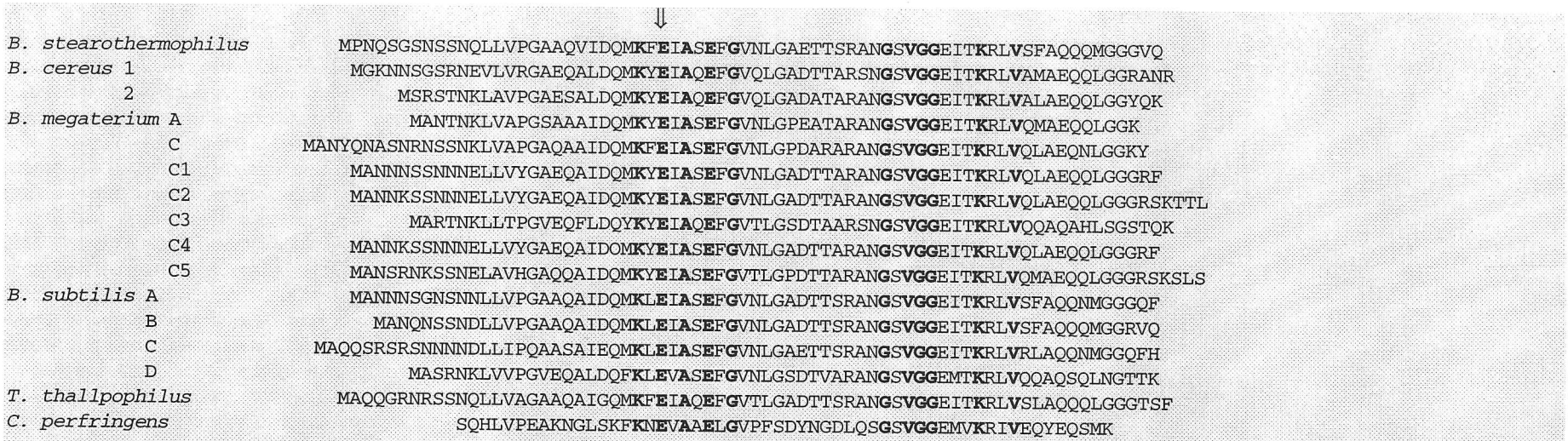


Figure 8.21. Comparison of amino acid sequences of α/β-type SASP from various species (Connors *et al.*, 1986; Connors & Setlow, 1985; Fliss *et al.*, 1986; Fliss & Setlow 1984; Fliss & Setlow, 1985; Granum *et al.*, 1987; Loshon *et al.*, 1986; Setlow & Ozols, 1979; Setlow & Ozols, 1980). Adapted from Setlow, 1988. The arrow (⇓) above the top sequence denotes the site of cleavage by the SASP-specific protease. Amino-acids conserved in all known SASP are marked in bold.

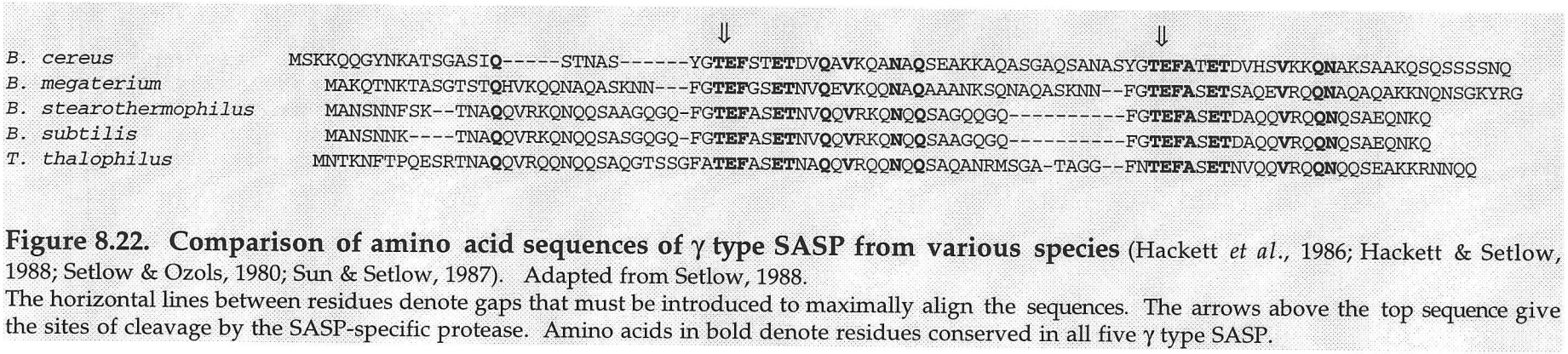


Figure 8.22. Comparison of amino acid sequences of γ type SASP from various species (Hackett *et al.*, 1986; Hackett & Setlow, 1988; Setlow & Ozols, 1980; Sun & Setlow, 1987). Adapted from Setlow, 1988.
The horizontal lines between residues denote gaps that must be introduced to maximally align the sequences. The arrows above the top sequence give the sites of cleavage by the SASP-specific protease. Amino acids in bold denote residues conserved in all five γ type SASP.

UV resistance below that of vegetative cells and well below that of wild type spores (Mason & Setlow, 1986) but otherwise grew and sporulated with kinetics indistinguishable from those in the wild-type strain (Hackett & Setlow, 1988; Mason & Setlow, 1986). These SASP $\alpha^-\beta^-$ mutants were also somewhat more heat sensitive than wild-type spores (Hackett & Setlow, 1988; Mason & Setlow, 1986)

It appears likely that the highly conserved region in the carboxyl-terminal half of these proteins is involved in DNA binding (Tovar-Rojo & Setlow, 1991). α/β-type SASP are bound to spore DNA *in vivo* and are double-stranded DNA-binding proteins *in vitro* (Francesconi *et al.*, 1988; Nicholson *et al.*, 1990; Setlow, 1992). The affinity between these proteins and DNA is not great, but their concentration *in vivo* is high, and there are sufficient amounts in spores to saturate the DNA (Nicholson *et al.*, 1990; Setlow, 1992; Setlow, 1988). Binding of these proteins alters the DNA's chemical and enzymatic reactivity as well as its UV photochemistry and is a key factor in spore DNA resistance to dessication, heat, oxidising agents and UV radiation (Setlow, 1995).

8.6.2 Nucleotide Sequence of a *Bacillus thuringiensis* SASP Gene

The nucleotide sequence of the 9890 bp lying furthest upstream of *am1* is presented in Figure 8.3. As has already been described, analysis of this nucleotide sequence revealed the presence of an open reading frame (1439-1659 bp) encoding a 74 amino acid protein OrfE with significant homology to α/β-type SASP from a wide range of sporulating bacteria.

Figure 8.23 shows the percentage similarity between OrfE and the α/β-type SASP from the highest scoring *Bacillus* species to come from a BLAST database search. The top five highest scoring matches in this search were all from *B. megaterium*, the greatest similarity being 60.3% with the C5 protein of *B. megaterium*. A multiple sequence alignment between OrfE and the same proteins is shown in Figure 8.24. From this alignment it is clear that although there are sequence differences between the proteins they share several important common features that are conserved in all SASP so far identified. In particular the two regions of amino acid sequence that are highly conserved in all other α/β-type SASP can be seen to be also present in OrfE. In both the region towards the amino terminal end that acts as a recognition and cleavage site for GPR (marked with an arrow) and the conserved region towards the carboxyl end of the protein believed to be responsible for DNA binding, OrfE shares all of the most highly conserved residues (marked in bold on figure). OrfE also lacks cysteine and tryptophan in common with all other α/β-type SASPs except those from *C. perfringens* some of which contain cysteine (Cabrera-Martinez & Setlow, 1991).

The coding region is preceded by a sequence with strong complementarity to the 3' end of Gram-positive bacterial 16S rRNA which is probably a ribosome binding site (underlined in Figure 8.3). The translation initiation codon is ATG and the termination

Percentage Similarity

	OrfE	BM C5	BS C	BC 1	
OrfE		60.3	58.3	55.7	OrfE
BM C5	35.2		56.9	55.7	BM C5
BS C	40.3	37.7		55.7	BS C
BC 1	41.2	30.0	40.3		BC 1
	OrfE	BM C5	BS C	BC 1	

Percentage Divergence

Figure 8.23 Protein sequence relationships between OrfE and other Bacillus α/β-type SASP using a Clustal PAM250 residue weight table.
BM C5 is from *Bacillus Megaterium* (Swiss/Prott accession P04835); BS C is from *Bacillus subtilis* (Swiss/Prott accession P02960); BC 1 is from *Bacillus cereus* (Swiss/Prott accession P06551).

```
                                                    ⇓
OrfE                MNIQRNESKSSTNEIFISAAASAIEQMKYEIARELGVTLGPDTSS
B. megaterium C5    MA---NSRNKSSNELAVHGAQQAIDQMKYEIASEFGVTLGPDTTA
B. subtilis C       MAQQSRSRSNNNNDLLIPQAASAIEQMKLEIASEFGVQLGAETTS
B. cereus 1         MG---KNNSGSRNEVLVRGAEQALDQMKYEIAQEFGVQLGADTTA
                    *          . *.  .  *  * .*** *** * ** ** .*..

OrfE                RANGSVGGEITKRLVRMAEEQLTGQYR--LH
B. megaterium C5    RANGSVGGEITKRLVQMAEQQLGGGRSKSLS
B. subtilis C       RANGSVGGEITKRLVRLAQQNMGGQF----H
B. cereus 1         RSNGSVGGEITKRLVAMAEQQLGG---RANR
                    *.************  *     *
```

Figure 8.24 Clustal protein sequence alignment of α/β-type SASP and the deduced sequence of the putative OrfB from *Bacillus thuringiensis* ssp. *fukuokaensis*.
See Figure 8.23 for accession numbers. Residues conserved in all four sequences are marked with an asterix and similar residues are marked with a full stop. Bold residues denote those conserved in all α/β-type SASP so far studied. The arrow above the top sequence gives the site of cleavage by the SASP-specific protease.

codon is TAA. Downstream of the translation codon is a region of dyad symmetry (underlined with double arrows in Figure 8.3) ending in an AT-rich region which may be a transcription terminator and is characteristic of SASP genes. The symmetry in this region is not perfect as it is downstream of many SASP genes with 2 residues out of the 20 not finding their complement in the opposing sequence. However, the symmetry is perfect for the first 16 residues of this sequence.

Another feature in common with other α/β-type SASP genes and indeed several sporulation specific genes including Bt *cry* genes (Wong *et al.*, 1983) and the sporulation gene *spoVG* (Banner *et al.*, 1983; Moran *et al.*, 1981) is an extended AT-rich region. This region has been shown to be essential for transcription of the *spoVG* gene (Banner *et al.*, 1983). For *orfE* this AT-rich region begins 36 bases upstream from the transcription start site and encompasses about 50 base pairs.

orfE has all these features in common with other previously cloned α/β-type SASP genes, what is unusual is its probable location on a plasmid (alongside a crystal protein gene that is almost always plasmid located). At first sight it seems strange that a gene needed for such an essential function of the cell should be encoded on plasmid DNA rather than associated with other genes involved in sporulation on the chromosome. However, the distribution of SASP genes has always been thought to be rather unusual and this is perhaps just another instance. For example *B. subtilis* α/β-type SASP genes are scattered around the chromosome, several in regions devoid of other known sporulation loci (Connors *et al.*, 1986). No satisfactory conclusion has ever been reached about how the many different α/β-type SASP genes arose and became scattered in such a manner on the chromosome. Whatever the mechanism it is not unlikely that SASP genes or copies of SASP genes could also have been transferred to plasmid DNA during the process. An obvious possibility is some duplicative transposition-like function. However, analysis of the sequences of many α/β-type SASP genes had so far revealed no transposon-like sequences. The identification of this α/β-type SASP gene on the 17A plasmid in a region possibly associated with transposase activity is therefore particularly interesting.

No work has been carried out to determine if *orfE* is expressed in Bt ssp *fukuokaensis* 17A. However, as is discussed above, shared similarities in upstream and downstream sequence as well as the high level of similarity with other SASP that are known to be expressed implies that the gene may well be expressed. Previous experiments have shown that there is no problem with SASP genes expression from introduced plasmids. α/β-type SASP have been introduced on plasmids to successfully cure UV-sensitive phenotypes of SASP- $\alpha^-\beta^-$ spores (Mason & Setlow, 1987), this had the same effect as integrating the same genes onto the chromosome.

8.6.3 Conclusions

SASPs capture many of the important ecogenic and developmental genetic events that produced the taxal diversity found within the *Clostridium* and *Bacillus* genera (Leighton, personal communication). SASP can therefore be exploited as 'molecular chronometers' to uncover the evolutionary history of these spore forming bacteria. Such a project is now being carried out at the Centre for Antimicrobial Discovery and Diagnostics, University of California. Bearing this in mind, it is therefore surprising that sequence analysis of this particular SASP gene from Bt should show it to be most closely related to SASP genes from *B. megaterium.* As is discussed in Chapter 1, *B. cereus* is thought to be the closest relative to Bt and one could have expected that this would be reflected in the phylogeny of their SASP genes. The discrepancy may be due to the unusual location of the Bt SASP gene on a plasmid. Perhaps this plasmid borne SASP is only present in certain *Bacillus* taxa.

The lack of any published Bt SASP sequences makes it impossible to draw any further conclusions. However, it is possible that in future, the phylogenetic study of Bt SASPs could be used in conjunction with current phylogenetic studies of Bt crystal protein genes to learn more about the evolutionary history of this organism.

Nine

Chapter 9

Conclusions and Proposals for Future Work

This project aimed to identify new, potentially diptericidal proteins. However, the discovery of a novel Cry10A-type protein and its association with mobile genetic elements related to those from other Bt strains harbouring similar genes, raised interesting phylogenetic considerations. When the *cry10A* gene was first identified in Bti, along with part of an *orf2* sequence that appeared to encode the C-terminal region (Thorne *et al.*, 1996), it was thought to be little more than a damaged *cry4A/cry4B* type gene. Since then the discovery that both this *cry10A* gene and the *ORF2* sequence are expressed as components of the Bti crystal (Pucell, 1997), and the identification of two more Cry proteins with similar gene arrangements (Cry19A [Rosso & Delécluse, 1997] and AM1), suggests that they may play a more interesting role both in toxicity and in the evolutionary development of the *cry* gene family.

A recent review on the phylogenetic relationships of Bt δ-endotoxins (Bravo, 1997) suggested that rather than testing each new protein against a growing series of organisms, phylogeny could be used to identify potential activities. Knowledge of the evolutionary relationships of the different Cry toxin domains could also be used to direct strategies to create novel chimeric toxins with different specificities (Bravo, 1997). With this in mind, a detailed phylogenetic analysis was carried out using *am1* and closely related *cry* genes.

Instead of analysing domains I, II and III as Bravo had done (Bravo, 1997), the N and C-terminal regions of Cry4A and Cry4B corresponding to ORF1 and ORF2 of the "two-gene" Cry10A, Cry19A and AM1 proteins were independently analysed. This information, along with expression and toxicity data raises a number of interesting points for future research.

1). AM1 toxicity assays

In this Thesis, AM1 spore/crystal mix was found to be non-toxic to *Aedes aegypti* at a concentration of 300 µg/ml. Rosso & Delécluse (1997) also found Cry19A to be non-toxic to this mosquito species, however it was toxic to both *Aedes stephensi* and *Culex pipiens*. The similarity between Cry19A and AM1 (nearly 50% [see Chapter 7] indicates that it would be worth testing both AM1 and 17A inclusions against both of these mosquito species.

The results of a preliminary bioassay indicated that AM1 and AM2 are at least partially responsible for the toxicity of the wild-type strain. Further bioassays should be

carried out to determine the relative contribution of these two proteins to the observed toxicity.

2). Study of *am1* and *am2* expression and toxicity

am1 does not appear to be expressed under its own promoter from the clone pSVP27AM1- which contains only part of the *am2* gene. A clone containing the entire *am1* and *am2* genes has been constructed (pSVP27AM1/2+), however the orientation of the insert in this construct places both genes under the control of the *cyt1A* promoter. Reversal of the insert orientation in this clone could determine if the presence of the complete *am2* gene is sufficient to allow transcription from the native promoter region. Chapter 6 describes how PCR could be used to generate restriction sites in pSVP27AM1/2+ or similarly in pSVP27AM1/2- to create clones containing isolated *am1* or *am2* genes. This library of clones could then be used in expression and toxicity studies to investigate the contibution of each component.

Rosso and Delécluse (1997) reported that Cry19A inclusions were less toxic to *C. pipiens* and *A. stephensi* than inclusions containing both Cry19A and ORF2. This suggested that either Cry19A and ORF2 are both toxic and could act synergistically, or that Cry19A is the only toxic polypeptide and ORF2 interacts with Cry19A at a transcriptional or post-transcriptional level, increasing the toxicity. The second hypothesis was thought to be more likely, as Cry19A was found to be poorly expressed on its own and appeared degraded upon SDS-PAGE analysis. More Cry19A also appeared to be produced when co-expressed with ORF2.

The presence of the *orf2* genes of Cry10A, Cry19A and AM1 may stabilise the mRNA, or the ORF2 may act as a chaperone to increase the stability of the ORF1 protein. The carboxy-terminal region of the 130-kDa Cry4B, similar to the ORF2 proteins, is not essential for its toxicity. In deletion analysis only the amino-terminal fragment (showing significant similarity to Cry10A, Cry19A and AM1 [see Chapter 7]) retains full activity (Chungjatupornchai *et al.*, 1988; Delécluse *et al.*, 1988; Pao-Intara *et al.*, 1988) suggesting that the C-terminal region is involved in stability rather than toxicity.

3). Generation of chimeric proteins

The cloned genes of *cry4A, cry4B, cry10A, cry19A* and *am1* along with the *orf2* region of *cry19A* and *am2* provide a library of N and C-terminal regions of very similar proteins with different specificities and potencies with which to generate chimeric proteins. If the entire *orf2* region of *cry10A* was also cloned this would enable additional combinations to be constructed.

Chapter 6 describes an attempt to create an AM1/AM2 chimera. It also suggests a method by which this genetic arrangement could be constructed by removal of the small

intergenic region between *am1* and *am2* to leave one continuous open reading frame. An investigation into the expression and toxicity of this protein compared with independently expressed *am1* and *am2* gene products may shed some light on the effect of the different gene arrangements on expression and toxicity, particularly if the reverse process is carried out for either *cry4A* or *cry4B*.

The hypothesis that the presence of a Cry4-type protein as either a two-gene or single-gene arrangement may be dependent on the protease activity in the susceptible insect gut is discussed in Chapter 5. The construction of the above clones could be used to test this hypothesis. It is possible that construction of the chimeric proteins may disrupt C-terminal activation. However, study of the alignment of AM1/2 with the deduced crystal protein structures of Cry1Aa and Cry3A in Chapter 5 (Figure 5.17) indicates that the C-terminal end of domain III is some distance upstream of the proposed mutation site and therefore the protease recognition site should not be affected.

Construction of the chimeric Cry10A/Cry4B protein (Zhang, personal communication) has already shown how toxicity can be affected by the attachment of a C-terminal region to an ORF1 protein. Cry10A was found to be non-toxic to *A. aegypti*, however, the Cry10A/4B chimera was weakly mosquitocidal (Purcell, 1997). The construction of such chimeras should reveal the importance of the C-terminal region in different insect guts. It may also generate novel proteins with enhanced toxicities to certain diptera species and create proteins which, when used in conjunction with existing proteins could delay the onset of resistance.

Detailed mutational analysis has been carried out by several workers on domains I, II and III of activated toxins (Chapter 1) to determine their roles in toxicity, however, very little attention has been paid to the C-terminal region. The high degree of sequence conservation between the C-terminal regions of otherwise more divergent proteins (Chapter 5, Figures 5.12a and 5.12b) strongly suggests that these regions play an important role in toxin function and should be thoroughly investigated.

4. Identification and analysis of mobile genetic elements.

The presence of such elements in the vicinity of δ-endotoxin genes and their possible role in the dissemination and evolution of the Cry gene family has been discussed. Further investigation into the DNA surrounding new and existing genes is likely to reveal additional transposable elements. A thorough phylogenetic analysis of all existing elements (such as has been carried out by Bravo (1997) on the *cry* gene family) may reveal more evidence as to the evolutionary history of the diverse family of Cry proteins.

This Thesis identified putative insertion sequences on either side of the "two gene" *am1, am2* arrangement and a putative integrase/recombinase that could together form a novel Class II transposon. As discussed in Chapter 8, further work should be carried out to

obtain a clone containing the sequence further downstream of the *am1* gene to determine if regions for site-specific resolution and terminal inverted repeats are present.

5 . Co-expression of Cry20A and Orf1 from Bt ssp. *fukuokanesis* 84-I

As discussed in Chapter 7 the 90-kD protein of 84-I (Cry20A) has been cloned and sequenced (Gill, personal communication) but no stable product could be obtained on expression. An Orf1 protein believed to posses chaperone properties has also been cloned from this strain but the two genes are currently in different laboratories. Further work should be carried out to determine if co-expression of these two genes stabilises the Cry20A protein. It would also be interesting to co-express *cry20A* with *am2* to see what effect the ORF2 region has on Cry20A stability.

Results of the experiments outlined above may further elucidate the role of the C-terminal region in toxicity and stability of the crystal protein. They may also go some way to determining why the two such different gene arrangements exist for what are otherwise very similar proteins.

Ten

Chapter Ten

References

Adams, L.F., Visick, J.E. & Whiteley, H.R. (1989); A 20-kilodalton protein is required for efficient production of the *Bacillus thuringiensis* subsp. *israelensis* 27-kilodalton crystal protein in *Escherichia coli*. *J. Bacteriol.* **171**, 521-530.

Adams, L.F., Brown, K.L. & Whiteley, H.R. (1991); Molecular cloning and characterisation of two genes encoding sigma factors. *J. Bacteriol.* **173**, 3846-3854.

Adams, L.F., Mathews, S., Ohara, P., Peterson, A. & Gurtler, H. (1994); Elucidation of the mechanism of CryIIIA overproduction in a mutagenized strain of *Bacillus thuringiensis* var. *tenebrionis*. *Molecular Microbiol.* **14**, 381-389.

Adang, M.J. & Spence, K.D. (1982); Biochemical comparisons of the peritrophic membranes of the lepidopterans *Orygia pseudotsugata* and *Manduca sexta*. *Comp. Biochem. Physiol.* **73B**, 645-649.

Agaisse, H. & Lereclus, D. (1995); How does *Bacillus thuringiensis* produce so much insecticidal crystal protein. *J. Bacteriol.* **177**, 6027-6032.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. & Young, I.G. (1981); Sequence and organisation of the human mitochondrial genome. *Nature* **290**, 457- 465.

Andrews, R.E., Kanda, K. & Bulla, L.A. (1982); *In vitro* and *in vivo* synthesis of the parasporal crystal of *Bacillus thuringiensis*. In *Gene Regulation in Bacilli* (Ganesan, A.T., Chang, S. and Hoch, J.A., Eds.), p. 121; Acad. Press, New York, 1982.

Angsuthanasombat, C. & Panyim, S. (1989); Biosynthesis of 130-kDa mosquito larvicides in the cyanobacteria *Agmenellum quadruplicatum* PR-6. *Appl. Environ. Microbiol.* **55**, 2428-2430.

Angsuthanasombat, C., Chungjatupornchai, W., Kertbundit, S., Luxananil, P., Settasatien, C., Wilairat, P. & Panyim, S. (1987); Cloning and expression of 130-kDa mosquito-larvicidal delta-endotoxin gene of *Bacillus thuringiensis* var. *israelensis* in *Escherichia coli*. *Mol. Gen. Genet.* **208**, 384-389.

Angsuthanasombat, C., Crickmore, N. & Ellar, D.J. (1991); Cytotoxicity of a cloned *Bacillus thuringiensis* subsp. *israelensis* CryIVB toxin to an *Aedes aegypti* cell line. *FEMS Microbiol. Lett.* **83**, 273.

Angsuthanasombat, C., Crickmore, N. & Ellar, D.J. (1992); Comparison of *Bacillus thuringiensis* subsp. *israelensis* CryIVB cloned toxins reveals synergism *in vivo*. *FEMS Microbiol. Lett.* **94**, 63-68.

Angsuthanasombat, C., Crickmore, N. & Ellar, D.J. (1993); Effects on toxicity of eliminating a cleavage site in a predicted interhelical loop in *Bacillus thuringiensis* CryIVB δ-endotoxin. *FEMS Microbiol. Lett.* **111**, 255-262.

Angus, T.A. (1956); Extraction, purification, and properties of *Bacillus sotto* toxin. *Can. J. Microbiol.* **2**, 416-426.

Arantes, O. & Lereclus, D. (1991). Construction of cloning vectors for *Bacillus thuringiensis*. *Gene*. **108**, 115-119.

Aronson, A.I., Beckman, W. & Dunn, P. (1986); *Bacillus thuringiensis* and related insect pathogens. *Microbiol. Rev.* **50**, 1-24.

Aronson, A.I., Han, E.-S., McGaughey, W. & Johnson, D. (1991); The solubility of inclusion proteins from *Bacillus thuringiensis* is dependent upon protoxin composition and is a factor in toxicity to insects. *Appl. Environ. Microbiol.* **57**, 981-986.

Aronson, A.I., Wu, D. & Zhang, C.L. (1995); Mutagenesis of specificity and toxicity regions of a *Bacillus thuringiensis* protoxin gene. *J. Bacteriol.* **177**, 4059-4065.

Arthur, A. & Sherratt, D.J. (1979); Dissection of the transposition process: a transposon-encoded site-specific recombination system. *Mol. Gen. Genet.* **175**, 267-274.

Ash, C., Farrow, J.A.E., Wallbanks, S. & Collins, M.D. (1991); Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small subunit ribosomal RNA sequences. *Lett. Appl. Microbiol.* **13**, 202-206.

Ausubel, F.M. (1991); In *Current Protocols in Molecular Biology*. Wiley Interscience, New York.

Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Seguin, C., Tuffnell, P.S. & Barrell, B.G. (1984); DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* **310**, 207.

Banner C.D.B., Moran, C.P.Jr & Losick, R. (1983); Deletion analysis of a complex promoter for a development regulated gene from *Bacillus subtilis*. *J. Mol. Biol.* **168**, 351-365.

Barberis-Maino, L., Berger-Bachi, B., Weber, H., Beck, W.D. & Kayser, F.H. (1987); IS*431*, a staphylococcal insertion sequence-like element related to IS*26* from *Proteus vulgaris*. *Gene* **59**, 107-113.

Barloy, F., Delécluse, A., Nicolas, L. & Lecadet, M.-M. (1996); Cloning and expression of the first anaerobic toxin gene from *Clostridium bifermentans* subsp. *malaysia*, encoding a new mosquitocidal protein with homologies to *Bacillus thuringiensis*. *J. Bacteriol.* **178**, 3099-3105.

Barton, K.A., Whiteley, H.R. & Yang, N.S. (1987); *Bacillus thuringiensis* δ-endotoxin expressed in transgenic *Nixotiana tabacum* provides resistance to lepidopteran insects. *Plant Physiology* **85**, 1103-1109.

Baum, J.A. (1994); Tn*5401*, a new Class II Transposable element from *Bacillus thuringiensis*. *J. Bacteriol.* **176**, 2835-2845.

Baum, J.A. & Malvar, T. (1995); Regulation of insecticidal crystal protein production in *Bacillus thuringiensis*. *Mol. Microbiol.* **18**, 1-12

Baum, J.A., Coyle, D.M., Gilbert, M.P., Jany, C.S. & Gawron-Burke, C. (1990); Novel cloning vectors for *Bacillus thuringiensis*. *Appl. Environ. Microbiol.* **56**, 3420-3428.

Baum, J.A., Kakefuda, M. & Gawron-Burke, C. (1996); Engineering *Bacillus thuringiensis* bio-insecticides with an indigenous site-specific recombination system. *Appl. Environ. Microbiol.* **62**, 4367-4373.

Bechtel, D.B. & Bulla, L.A. (1976); Electron microscopic study of sporulation and parasporal crystal formation in *Bacillus thuringiensis*. *J. Bacteriol.* **127**, 1472-1481.

Becker, N. & Margalit, J. (1993); Use of *Bacillus thuringiensis israelensis* against mosquitoes and blackflies, p145-170. In P.F. Entwistle, P.F. Cory, M.J. Bailey, S. Higgs (ed.), *Bacillus thuringiensis, an Environmental Biopesticide: Theory and Practice*. J. Wiley & Sons, New York, N.Y.

Beegle, C.C. & Yamamoto, T. (1992); The history of *Bacillus thuringiensis berliner* - research and development. *Can. Entomol.* **124**, 587-616.

Berenbaum, M. (1980); Adaptive significance of midgut pH in larval lepidoptera. *Am. Naturalist* **115**, 138-146.

Berliner, E. (1915); Uber die schlaffsucht der mehlmottenraupe (*Ephestia kuhniella* zell) und ihren Erreger *Bacillus thuringiensis*, sp. *Angew. Entomol.* **2**, 29-56.

Berner, R., Rudin, W. & Hecker, H. (1983); Peritrophic membranes and protease activity in the midgut of the malarial mosquito, *Anopheles stephensi* (Liston) (Insecta: Diptera) under normal and experimental conditions. *Journal of Ultrastructural Research* **83**, 195-204.

Biel, S.W. & Berg, D.E. (1984); Mechanism of IS*1* transposition in *E. coli*: choice between simple insertion and co-integration. *Genetics* **108**, 319-330.

Bietlot, H.P.L., Vishnubhatha, I., Carey P.R., Pozsgay, M. & Kaplan, H. (1990); Characterisation of the cysteine residues and disulphide linkages in the protein crystal of *Bacillus thuringiensis*. *Biochem. J.* **267**, 309-315.

Birnboim, H.C. & Doly, J. (1979); A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* **7**, 1513-1523.

Bone, L.W. (1989); Activity of commercial *Bacillus thuringiensis* preparations against *Trichostrongylus colubriformis* and *Nippostrongylus brasiliensis*. *J. Invertebr. Pathol.* **53**, 276-277.

Bone, E.J. & Ellar, D.J. (1989); Transformation of *Bacillus thuringiensis* by electroporation; *FEMS Microbiol.* **58**, 171-178.

Bone, L.W., Bottjer, K.P. & Gill, S.S. (1988); Factors affecting the larvicidal activity of *Bacillus thuringiensis israelensis* toxin for *Trichostrongylus colubriformis* (Nematoda). *J. Invertebr. Pathol.* **52**, 102-107.

Borovsky, D. (1986); Proteolytic-enzymes and blood digestion in the mosquito, *Culex nigripalpus*. *Archives of Insect Biochemistry and Physiology* **3**, 147-160.

Bosch, D., Schipper, B., Van der Kliej, H., de Maagd, R.A. & Stiekema, W.J. (1994); Recombinant *Bacillus thuringiensis* crystal proteins with new properties; possibilities for resistance management. *Bio/Technology* **12**, 915-919.

Bourgouin, C., Delécluse, A., Ribier, J., Klier, A. & Rapoport, G. (1988); A *Bacillus thuringiensis* subsp. *israelensis* gene encoding a 125-kilodalton larvicidal polypeptide is associated with inverted repeat sequence. *J. Bacteriol.* **170**, 3575-3583.

Brandt, C.R., Adang, M.J. & Spence, K.D. (1978); The peritrophic membrane: Ultrastructural analysis and function as a mechanical barrier to microbial infection in *Orgyia pseudotsugata*. *J. Invertebr. Pathol.* **32**, 12-24.

Bravo, A. (1997); Phylogenetic relationships of *Bacillus thuringiensis* delta-endotoxin family proteins and their functional domains. *J. Bacteriol.* **9**, 2793-2801.

Briggs, J.D. (1986); Pioneering and advanced phases of commercial use of *Bacillus thuringiensis* in North America. pp. 25-35 in Krieg, A & Huger, A.M. (Eds.), *Mitt. Biol. Bundestanst. Land Forstwirtsch. Berl. Dahlem.* Vol. **233**. Paul Parcet, Berlin.

Brizzard, B.L. & Whiteley, H.R. (1988); Nucleotide-sequence of an additional crystal protein gene cloned from *Bacillus thuringiensis* subsp. *thuringiensis*. *Nucleic Acids Research* **16**, 2723-2724.

Brizzard, B.L., Schnepf, H.E. & Kronstad, J.W. (1991); Expression of the *cryIB* crystal protein gene of *Bacillus thuringiensis*. *Mol. Gen. Genet.* **231**, 59- 64.

Brousseau, R. & Masson, L. (1988); *Bacillus thuringiensis* insecticidal crystal toxins: gene structure and mode of action. *Biotech. Adv.* **6**, 697-724.

Brown, K.L. (1993); Transcriptional regulation of the *Bacillus thuringiensis* subsp. *thompsoni* crystal protein operon. *J. Bacteriol.* **175**, 7951-7957.

Brown, K.L. & Whiteley, H.R. (1988); Isolation of a *Bacillus thuringiensis* RNA polymerase capable of transcribing crystal protein genes. *Proc. Natl. Acad. Sci. USA*. **85**, 4166-4170.

Brown, K.L. & Whiteley, H.R. (1990); Isolation of the second *Bacillus thuringiensis* RNA polymerase that transcribes from a crystal protein gene promoter. *J. Bacteriol.* **172**, 6682-6688.

Brown, K.L. & Whiteley, H.R. (1992); Molecular characterization of two novel crystal protein genes from *Bacillus thuringiensis* subsp. *thompsoni*. *J. Bacteriol.* **174**, 549-557.

Bulla, L.A., Kramer, K.J. & Davidson, L.I. (1977); Characterisation of the entomocidal parasporal crystals of *Bacillus thuringiensis*. *J. Bacteriol.* **130**, 375-383.

Bulla, L.A., Kramer, K.J., Cox, D.J., Jones, B.L., Davidson, L.I. & Lookhart, G.L. (1981); Purification and characterization of the entomocidal protoxin of *Bacillus thuringiensis*. *J. Biol. Chem.* **256**, 3000-3004.

Ceron, J., Ortiz, A., Quintero, R., Guereca, L & Bravo, A. (1995); Specific PCR primers directed to identify *cryI* and *cryIII* genes within a *Bacillus thuringiensis* strain collection. *Appl. Environ. Microbiol.* **61**, 3326-3831

Burges, H.D. & Hurst, J.A. (1977); Ecology of *Bacillus thuringiensis* in storage moths. *J. Invertebr. Pathol.* **30**, 131-139.

Burton, S.L., Derbyshire, D., Ellar, D.J., Li, J. & Murzin, A. (1998); Triple mutant in domain III completely eliminated GalNAc binding. In *The VIIth International Colloquium on International Pathology and Microbial Control*; Sapporo, Japan.

Cabrera-Martinez, R.M. & Setlow, P. (1991); Cloning and nucleotide sequence of three genes coding for small, acid-soluble proteins of *Clostridium perfringens* spores. *FEMS Microbiol. Lett.* **77**, 127-131.

Carlson, C.R., Caugant, D.A. & Kolstø, A.-B. (1994); Genotypic diversity among *Bacillus cereus* and *Bacillus thuringiensis* strains. *Appl. Environ. Microbiol.* **60**, 1719-1725.

Carlton, B.C., Gawron-Burke, C. & Johnson, T.B. (1990); Exploiting the genetic diversity of *Bacillus thuringiensis* for the creation of new bio-insecticides. In *Fifth International Colloquium on Invertebrate Pathology and Microbial Control*: 1990, pp. 18-22; Society for Invertebrate Pathology.

Carlton, B.C. & Gonzalez, J.M. (1985); The genetics and molecular biology of *Bacillus thuringiensis*. In *The Molecular Biology of the Bacilli*, Vol. **2** (Dubanu, D.A., Ed.), pp. 211-249, Academic Press Inc., New York.

Carroll, J. & Ellar, D.J. (1993); An analysis of *Bacillus thuringiensis* δ-endotoxin action on insect-midgut-membrane permeability using a light-scattering assay. *Eur. J. Biochem.* **214**, 771-778.

Carroll, J., Convents, D., Van Damme, J., Boets, A., Van Rie, J. & Ellar, D.J. (1997); Intramolecular proteolytic cleavage of *Bacillus thuringiensis* Cry3A delta-endotoxin may facilitate its coleopteran toxicity. *J. Invertebr. Pathol.* **70**, 41-49.Ceron, J., Ortiz, A., Quintero, R., Guereca, L. & Bravo, A. (1995); Specific PCR primers directed to identify *cryI* and *cryII* genes within a *Bacillus thuringiensis* strain collection. *Appl. Environ. Microbiol.* **61**, 3826-3831.

Chambers, J.A., Jelen, A., Gilbert, M.P., Jany, C.S., Johnson, T.B. & Gawron-Burke, C. (1991); Isolation and characterization of a novel insecticidal crystal protein gene from *Bacillus thuringiensis* subsp. *aizawai*. *J. Bacteriol.* **173**, 3966-3976.

Chambers, C.E., Bone, E. & Ellar, D.J. (1997); Protease activity of three lepidopteran species during larval development. In *30th Annual Meeting of the Society for Invertebrate Pathology*, Banff, Canada.

Chang, C., Yu, Y.M., Dai, S.M., Law, S.K. & Gill, S.S. (1993); High level CryIVD and CytA gene expression in *Bacillus thuringiensis* does not require the 20-kilodalton protein and the co-expressed gene products are synergistic in their toxicity to mosquitoes. *Appl. Environ. Microbiol.* **59**, 815-821.

Chanpaisang, J., Attathom, T. & Thaveechai, N. (1994); Identification of local strains of *Bacillus thuringiensis* and their efficacy tests. In *Abstracts of the 6th International Colloquium on Invertebrate Pathology and Microbial Control,* Montpellier, p. 359.

Chassy, B.M. & Flickinger, J.L. (1987); Transformation of *Lactobacillus casei* by electroporation. *FEMS Microbiol. Lett.* **44**, 173-177.

Chen, X.J., Lee, M.K. & Dean, D.H (1993); Site-directed mutations in a highly conserved region of *Bacillus thuringiensis* δ-endotoxin affect inhibition of short-circuit current across *Bombyx mori* midguts. *Proc. Natl. Acad. Sci. USA.* **90**, 9041-9045.

Cheng, Y.S.E. & Aronson, A.I. (1977); *Proc. Natl. Acad. Sci. USA.* **74**, 1254-1258.

Chestukhina, G.C., Kostina, L.I., Mikhailova, A.L., Tyurin, S.A., Klepikova, F.S. & Stepanov, V.M. (1982); The main features of *Bacillus thuringiensis* δ-endotoxin molecular structure. *Arch. Microbiol.* **132**, 159-162.

Chilcott, C.N. & Ellar, D.J. (1988); Comparative toxicity of *Bacillus thuringiensis* var. *israelensis* crystal proteins *in vivo* and *in vitro*. *J. Gen. Microbiol.* **134**, 2551-2558.

Choma, C.T., Surewicz, W.K. & Kaplan, H. (1991); The toxic moiety of the *Bacillus thuringiensis* protoxin undergoes a conformational change upon activation. *Biochem. Biophys. Res. Commun.* **179**, 933-938.

Chungjatupornchia, W., Höfte, H., Seurinck, J., Angsuthanasombat, C. & Vaeck, M. (1988); Common features of *Bacillus thuringiensis* toxins specific for diptera and lepidoptera; *Eur. J. Biochem.* **173**, 9-16.

Church, B.D. & Halvorson, H. (1957); Intermediate metabolism of aerobic spores. Inactivation of glucose oxidation in spores of *Bacillus cereus* var *terminalis*. *J. Bacteriol.* **73**, 470-476.

Connors, M.J. & Setlow, P. (1985); Cloning of a gene for a small, acid-soluble spore protein from *Bacillus subtilis* and determination of its complete nucleotide sequence. *J. Bacteriol.* **161**, 333-339.

Connors, M.J., Howard, S. Hock, J. & Setlow, P. (1986); Determination of the chromosomal locations of four *Bacillus subtilis* genes which code for a family of small, acid-soluble spore proteins. *J. Bacteriol.* **166**, 412-416.

Couche, G.A., Pfannenstiel, M.A. & Nickerson, K.W. (1987); Structural disulphide bonds in the *Bacillus thuringiensis* subsp. *israelensis* protein crystal. *J. Bacteriol.* **169**, 3281-3288.

Crickmore, N., Bone, E.J. & Ellar, D.J. (1990); Genetic manipulation of *Bacillus thuringiensis*: towards an improved pesticide. *Aspects Appl. Biol.* **24**, 17-24.

Crickmore, N. & Ellar, D.J. (1992); Involvement of a possible chaperone in the efficient expression of a cloned CryIIA δ-endotoxin gene in *Bacillus thuringiensis*. *Mol. Microbiol.* **6**, 1533-1537.

Crickmore, N., Wheeler, V.C. & Ellar, D.J. (1994); Use of an operon fusion to induce expression and crystallization of a *Bacillus thuringiensis* δ-endotoxin encoded by a cryptic gene. *Mol. Gen. Genetics* **242**, .365-368.

Crickmore, N., Bone, E.J., Williams, J.A. & Ellar, D.J. (1995); Contribution of the individual components of the δ-endotoxin crystal to the mosquitocidal activity of *Bacillus thuringiensis* subsp. *israelensis*.. *FEMS Microbiol. Lett.* **131**, 249-254.

Crickmore, N., Zeigler, D.R., Feitelson, J., Schnepf, E., V an Rie, J., Lereclus, D., Baum, J. & Dean, D.H. (1998); Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. *Microbiol. Mol. Biol. Rev.* **62**, 807-813.

Cummings, C.E., Armstrong, G., Hodgman, T.C. & Ellar, D.J. (1994); Structural and functional studies of a synthetic peptide mimicking a proposed membrane inserting region of a *Bacillus thuringiensis* δ-endotoxin. *Mol. Membr. Biol.* **11**, 87-92.

Dadd, R.H. (1975); Alkalinity within the midgut of mosquito larvae with alkaline-active digestive enzyme. *J. Insect. Physiol.* **21**, 1847-1853.

Dalrymple, B., Caspers, P. & Arber, W. (1984); Characterization of the mobile genetic element IS*30*. *Experimentia* **40**, 622

Damgaard, P.H. (1995); Diarrheal enterotoxin production by strains of *Bacillus thuringiensis* isolated from commercial *Bacillus thuringiensis*-based insecticides. *FEMS Immunol. Med. Microbiol.* **12**, 245-249.

Damgaard, P.H., Hansen, B.M., Pederson, J.C. & Eilenberg, J. (1994); Naturally occurring *Bacillus thuringiensis* strains on foliage, in soil and in insects. In *Abstracts of the 6th International Colloquium on Invertebrate Pathology and Microbial Control*, Montpellier, p37.

Dastidar, P. G. & Nickerson, K. W. (1979); Interchain crosslinks in the entomocidal *Bacillus thuringiensis* protein crystal. *FEBS Lett.* **108**, 411-414.

de Barjac, H. (1978); Une nouvelle variete de *Bacillus thuringiensis* tres toxique pour les Moustiques: *B. thuringiensis var. israelensis*, serotype 14; *Comptes Rendus Academie Sciences Paris*, serie D **268**, 797-800.

de Barjac, H. & Bonnefoi, A. (1968); A classification of strains of *Bacillus thuringiensis* Berliner with a key to their differentiation. *J. Invertebr. Pathol.* **11**, 335-347.

de Barjac, H. & Frachon, (1990); Classification of *Bacillus thuringiensis* strains. *Entomophaga* **35**, 233.

Debro, L., Fitzjames, P.C. & Aronson, A (1986); Different parasporal inclusions are produced by *Bacillus thuringiensis* ssp. *finitimus*. *J. Bacteriol.* **165**, 258-268.

Delécluse, A., Bougouin, C., Klier, A. & Rapoport, G. (1989); Nucleotide sequence and characterisation of a new insertion element, IS*240*, from *Bacillus thuringiensis israelensis*. *Plasmid* **21**, 71-78.

Delécluse, A., Bourgouin, C., Menou, G., Lereclus, D., Klier, A. & Rapoport, G. (1990); IS*240* associated with the *cryIVA* gene from *Bacillus thuringiensis israelensis* belongs to a family of Gram(+) and Gram(-) IS elements; In *Genetics and Biotechnology of Bacilli* Vol. **3** (Zukowski, M.M., Ganesan, A.T. and Hoch, J.A., Eds.), pp. 181-190; Academic Press, San Diego, CA.

Delécluse, A., Rosso, M.L. & Ragni, A. (1995); Cloning and expression of a novel toxin gene from *Bacillus thuringiensis* subsp. *jegathesan* encoding a highly mosquitocidal protein. *Appl. Environ. Microbiol.* **61,** 4230-4235.

DeLucca, A.J., Palmgren, M.S. & Ciegler, A. (1982); *Bacillus thuringiensis* in grain elevator dusts. *Can. J. Microbiol.* **28**, 452-456.

Dervyn, E., Poncet, S., Klier, A. & Rapoport, G. (1995); Transcriptional regulation of *Bacillus thuringiensis* subsp. *israelensis cryIVD* operon. *J. Bacteriol.* **177**, 2283-2291.

Dignam, S.S. & Setlow, P. (1980); *In vivo* and *in vitro* synthesis of the spore-specific proteins A and C of *Bacillus megaterium*. *J. Biological Chem.* **255**, 8417-8423.

Dodd, I.B. & Egan, J.B. (1987); Systematic method for the detection of potential gamma-cro-like DNA-binding regions in proteins. *J. Mol. Biol.* **194**, 557-564.

Donovan, W.P., Dankocsik, C. & Gilbert, M.P. (1988a); Molecular characterization of a gene encoding a 72-kiolodalton mosquito-toxic crystal protein from *Bacillus thuringiensis* subsp. *israelensis*. *J. Bacteriol.* **170**, 4732-4738.

Donovan, W.P., Dankocsik, C.C., Gilbert, M.P., Gawron-Burke, M.C., Groat, R.G. & Carlton, B.C. (1988b); Amino acid sequence and entomocidal activity of the P2 crystal protein. *J. Biol. Chem.* **263**, 561-567. (Author's correction, **263**, 4740).

Donovan, W.P., Tan, Y., Jany, C.S. & González, Jr. J.M. (June 1994); U.S. patent 5,322,687.

Dow, J.A.T. (1984); Extremely high pH in biological systems: A model for carbonate transport. *Am. J. Physiol.* **246**, 633-635.

Dower, W.J., Miller, J.F. & Ragsdale, C.W. (1988); High efficiency transformation of *E. coli* by high voltage electroporation. *Nucl. Acids Res.* **16,** 6127-6145.

Drobniewski, F.A., Knowles, B.H. & Ellar, D.J. (1987); Nonspecific ionic effects on the cytolytic and haemolytic properties of *Bacillus thuringiensis* toxins. *Curr. Microbiol.* **15**, 295-299.

Drobniewski, F.A. & Ellar, D.J. (1988); Toxin-membrane interactions of *Bacillus thuringiensis* δ-endotoxins. *Biochem. Soc. Trans.* **16**, 39-40.

Drobniewski, F.A. & Ellar, D.J. (1989); Purification and properties of a 28-kilodalton haemolytic and mosquitocidal protein toxin of *Bacillus thuringiensis* subsp. *darmstadiensis* 73-E10-2. *J. Bacteriol.* **171**, 3060-3067.

Du, C. & Nickerson, K.W. (1996); *Bacillus thuringiensis* HD-73 spores have surface-localized Cry1Ac toxin: physiological and pathogenic consequences. *Appl. Environ, Microbiol.* **62**, 3722-3726.

Du, C., Martin, P.A.W. & Nickerson, K.W. (1994); Comparison of disulphide contents and solubility at alkaline pH of insecticidal and non-insecticidal *Bacillus thuringiensis* protein crystals. *Appl. Environ. Microbiol.* **60**, 3847-3853.

Dulmage, H.T. & Cooperators. (1981); Insecticidal activity of isolates of *Bacillus thuringiensis* and their potential for pest control; In *Microbial Control of Pests and Plant Diseases 1970-1980* (Burges, H.D., Ed.), pp. 193-222; Academic Press Inc., London.

Dulmage, H.T. & Aizawa, K. (1982); Distribution of *Bacillus thuringiensis* in nature; In *Microbial and Viral Pesticides* (Kurstak E., Ed.), pp. 209-237; Marcel Dekker, Inc., New York.

Dunn, M.G. (1996); Molecular characterisation of a *Bacillus thuringiensis* genetic locus. PhD Thesis, University of Cambridge.

Dunn, M.G. & Ellar, D.J. (1997); Identification of two sequence elements associated with the gene encoding the 24-kDa crystalline component in *Bacillus thuringiensis* ssp. *fukuokaensis*. An example of transposable element archaeology. *Plasmid* **37**, 205-215.

Dure, L.S., III. (1973); in *Developmental regulation*: Aspects of Cell Differentiation. (Coward, S.J., eds) pp.23-48, Academic Press, New York.

Earp, D.J., Ward, E.S. & Ellar, D.J. (1987); Investigation of possible homologies between crystal proteins of three mosquitocidal strains of *Bacillus thuringiensis*. *FEMS Microbiol. Lett.* **42**, 195-199.

Edlund, T., Siden, I. & Boman, H.G. (1976); Evidence for two immune inhibitors from *Bacillus thuringiensis* interfering with the humoral defense system of saturniid pupae. *Infect. Immun.* **14**, 934-941.

Edwards, D.L., Payne, J., & Soares, G.G. (1990); U.S. patent No. 4,948,734.

Ellar, D.J. (1990); Pathogenicity determinants of entomopathogenic bacteria; In *Fifth International Colloquium on Invertebrate Pathology and Microbial Control: 1990*, pp. 298-302; Society for Invertebrate Pathology.

Ellar, D.J., Knowles, B.H., Carroll, J., Horsnell, J., Haider, M.Z., Ahmad, W., Nicholls, C.N., Armstrong, G. & Hodgman, C. (1990); Genetics and biochemical studies of the mechanism of action of *Bacillus thuringiensis* entomocidal δ-endotoxins; In *Fourth European Workshop on Bacterial Toxins* (Rappouli, R., Alouf, J., Freer, J., Fehrenbach, F., Wadstrom, T. and Witholt, B., Eds.), pp. 499-506; Gustav Fischer Verlag, Stuttgart and New York.

English, L.H., Readdy, T.L. & Bastian, A.E. (1991); Delta-endotoxin induced leakage of $^{86}Rb^{+}$ - K^{+} and H_2O from phospholipid vesicles is catalyzed by reconstituted midgut membrane. *Insect Biochem.* **21**, 177-184.

Estruch, J.J., Warren, G.W., Mullins, M.A., Nye, G.J., Craig, J.A. & Koziel, M.G. (1996); Vip3A, a novel *Bacillus thuringiensis* vegetative insecticidal protein with a wide spectrum of activities against lepidopteran insects. *Proc. Natl. Acad. Sci. USA* **93**, 5389-5394.

Fast, P.G. (1966); A comparative study of the phospholipids and fatty acids of some insects. *Lipids* **1**, 209-215.

Faust, R.M., Travers, R.S. & Hallam, G.M. (1974); Preliminary investigations on the molecular modes of action of the δ-endotoxin produced by *Bacillus thuringiensis* var. *alesti*. *J. Invertebr. Pathol.* **23**, 259-261.

Federici, B.A., Lüthy, P. & Ibarra, J.E. (1990); Parasporal body of *Bacillus thuringiensis* subspecies *israelensis*: Structure, protein composition, and toxicity; In *Bacterial Control of Mosquitoes and Blackflies: Biochemistry, Genetics, and Applications of Bacillus thuringiensis and Bacillus sphaericus* (deBarjac, H. & Sutherland, D., Eds.), pp. 16-44; Rutgers University Press, New Brunswick.

Feitelson, J. S. (1993); The *Bacillus thuringiensis* Family Tree. In *Advanced Engineered Pesticides* (Kim, L., ed.), pp. 63-72. Marcel Dekker, Inc., New York.

Ferré, J., Escriche, B., Bel, Y. & Van Rie, J. (1995); Biochemistry and genetics of insect resistance to *Bacillus thuringiensis* insecticidal crystal proteins. *FEMS Microbiol. Lett.* **132**, 1-7.

Fliss, E.R. & Setlow, P. (1984); *Bacillus megaterium* spore protein C-3: nucleotide sequence of its gene and the amino acid sequence at its spore protease cleavage site. *Gene* **30**, 167-170.

Fliss, E.R & Setlow, P. (1985); Genes for *Bacillus megaterium* small, acid-soluble spore proteins, nucleotide sequence of two genes and their expression during sporulation. *Gene* **35**, 151-157.

Fliss, E.R., Loshon, C.A. & Setlow, P. (1986); Genes for *Bacillus megaterium* small, acid-soluble spore proteins, cloning and nucleotide sequence of nucleotide sequence of three additional genes from this multigene family. *J. Bacteriol.* **165**, 467-473.

Fox, J.L. (1995); EPA'S first commercial release is still pending. *Bio/Technology* **13**, 115-116.

Francesconi, S.C., Macalister, T.J., Setlow, B. & Setlow, P. (1988); Immunoelectron microscopic localization of small, acid-soluble spore proteins in sporulating cells of *Bacillus subtilis*. *J. Bacteriol.* **170**, .5963-5967.

Garczynski, S.F., Crim, J.W. & Adang, M.J. (1991); Identification of putative insect brush border membrane-binding molecules specific to *Bacillus thuringiensis* δ-endotoxin by protein blot analysis. *App. Environ. Microbiol.* **57**, 2816-2820.

Garduno, F., Thorne, L., Walfield, A.M. & Pollock, T.J. (1988); Structural relatedness between mosquitocidal endotoxins of *Bacillus thuringiensis* ssp. *israelensis*. *Appl. Environ. Microbiol.* **54**, 277-279.

Gawron-Burke, C. & Baum, J. (1991); Genetic manipulation of *Bacillus thuringiensis* insecticidal crystal protein genes in bacteria. In *Genetic Engineering* Vol. 13 (Setlow, J. K., ed.), pp. 237-263. Plenum Press, New York.

Gazit, E., Bach, D., Kerr, I. D., Sansom, M. S. P., Chejanovsky, N. & Shai, Y. (1994); The alpha-5 segment of *Bacillus thuringiensis* delta-endotoxin - *in vitro* activity, ion channel formation and molecular modeling. *Biochem. J.* **304**, 895-902.

Ge, A.Z., Rivers, D., Milne, R. & Dean, D.H. (1991); Functional domains of *Bacillus thuringiensis* insecticidal crystal proteins, refinement of *Heliothis virescens* and *Trichoplusia ni* specificity domains on CryIA(c). *J. Biol. Chem.* **266**, 17954-17958.

Georghiou, G. P., Wirth, M. C., Delécluse, A. & Klier, A. (1993); Potentiality for development of resistance to single vs. multiple toxins of *Bacillus* by mosquitoes. *VIIth International Conference on Bacillus,* Pasteur Institute, Paris, France.

Georghiou, G.P. & Wirth, M.C. (1997); Influence of exposure to single versus multiple toxins of *Bacillus thuringiensis* subsp. *israelensis* on development of resistance in the mosquito *Culex quinquefasciatus* (Diptera: Culicidae). *Appl. Environ. Microbiol.* **63**, 1095-1101.

Gill, S.S., Singh, G.J.P. & Hornung, J.M. (1987); Cell membrane interaction of *Bacillus thuringiensis* subsp. *israelensis* cytolytic toxins. *Infect. Immunol.* **55**, 1300-1308.

Gill, S.S., Cowles, E.A. & Pietrantonio, P.V. (1992); The mode of action of *Bacillus thuringiensis* endotoxins. *Ann. Rev. Entomol.* **37**, 615-636.

Gill, S.S., Cowles, E.A. & Francis, V. (1995); Identification, isolation, and cloning of a *Bacillus thuringiensis* CryIA(c) toxin-binding protein from the midgut of the lepidopteran insect *Heliothis virescens. J. Biol. Chem.* **270**, 27277-27282.

Gingrich, R.E., Allan, N. & Hopkins, D.E. (1974); *Bacillus thuringiensis:* Laboratory tests against four species of biting lice (Mallophaga: Trichodectidae). *J. Invertebr. Pathol.* **23**, 232-236.

Glatron, M.F. & Rapoport, G. (1972); Biosynthesis of the parasporal inclusion of *Bacillus thuringiensis*: half-life of its corresponding messenger RNA. *Biochemie* **54**, 1291-1301.

Gonzalez, J.M. & Carlton, B.C. (1982); In *Genetic Exchange: a Celebration and a New Generation* (Streips, U.N., Goodgal, S., Guild, W.R. and Wilson G., Eds.), pp. 89-95; Dekker, New York.

Gonzalez, J.M., Dulmage, H.T. & Carlton, B.C. (1981); Correlation between specific plasmids and δ-endotoxin production in *Bacillus thuringiensis. Plasmid* **5**, 351-365.

Gordon, R.E. (1977); Some taxonomic observations on the genus *Bacillus*. In *Biological Regulations of Vectors: The Saprophytic and Aerobic Bacteria and Fungi* (Briggs, J.B., Ed.). Pupl. NIH 77-1180, pp. 67-82. US Department of Health, Education and Welfare, Washington, USA.

Gordon, R.E., Haynes, W.C. & Pang, C.H.-N. (1973); The Genus *Bacillus*, Agricultural Handbook No. 427, US Department of Agriculture, Washington D.C.

Gould, F., Martinez-Ramirez, A., Anderson, A., Ferré, J., Silva, F. J. & Moar, W. J. (1992); Broad-spectrum resistance to *Bacillus thuringiensis* toxins in *Heliothis virescens. Proc. Natl. Acad. Sci. USA* **89**, 7986-7990.

Granum, P.E., Richardson, M., Blom, H. (1987); Isolation and amino acid sequence of an acid soluble protein from *Clostridium perfringens* spores. *FEMS Lett.* **42**, 225-230.

Granum, P.E., Pinnavaia, S.M. & Ellar, D.J. (1988); Comparison of the *in vivo* and *in vitro* activity of the δ-endotoxin of *Bacillus thuringiensis* var. *morrisoni* (HD-12) and two of its constituent proteins after cloning and expression in *Escherichia coli*. *Eur. J. Biochem.* **172**, 731-738.

Green, B.D., Battisti, L. & Thorne, C.B. (1989); Involvement of Tn*4430* in transfer of *Bacillus anthracis* plasmids mediated by *Bacillus thuringiensis* plasmids pXO12. *J. Bacteriol.* **171**, 104-113.

Griego, V.M., Moffet, D. & Spence, K.D. (1979); Inhibition of active K^+ transport in the Tobacco Hornworm (*Manduca Sexta*) midgut after ingestion of *Bacillus thuringiensis* endotoxin. *J. Insect Physiol.* **25**, 283-288.

Grochulski, P., Masson, L., Borisova, S., Pusztaicarey, M., Schwartz, J.L., Brousseau, R. & Cygler, M. (1995); *Bacillus thuringiensis* CryIA(a) insecticidal toxin crystal structure and channel formation. *J. Mol. Biol.* **254**, 447-464.

Hackett, R.H., Setlow, B. & Setlow, P. (1986); Cloning and nucleotide sequence of the *Bacillus megaterium* gene coding for small acid soluble spore protein B. *J. Bacteriol.* **168**, 1023-1025.

Hackett, R.H. & Setlow, P. (1987); Cloning, nucleotide sequencing and genetic mapping of the gene for small, acid-soluble spore protein-gamma of *Bacillus subtilis,. J. Bacteriol.* **169**, 1985-1992.

Hackett, R.H. & Setlow, P. (1988); Properties of spores of *Bacillus subtilis* which lack the major small, acid-soluble protein. *J. Bacteriol.* **170**, 1403-1404.

Haider, M.Z. & Ellar, D.J. (1987a); Characterization of the toxicity and cytopathic specificity of a cloned *Bacillus thuringiensis* crystal protein using insect cell culture. *Mol. Microbiol.* **1**, 59-66.

Haider, M.Z. & Ellar, D.J. (1987b); Analysis of the molecular basis of insecticidal specificity of *Bacillus thuringiensis* crystal δ-endotoxin. *Biochem. J.* **248**, 197-201.

Haider, M. Z., Knowles, B. H. & Ellar, D. J. (1986); Specificity of *Bacillus thuringiensis* var. *aizawai* insecticidal δ-endotoxin is determined by differential proteolytic processing of the protoxin by larval gut proteases. *Eur. J. Biochem.* **156**, 531-540.

Halling, S.M., Simons, R.W., Way, J.C., Walsh, R.B & Kleckner, N. (1982); DNA-sequence organisation of IS*10*-right of TN*10* and comparison with IS*10*-left. *Proc. Natl. Acad. Sci. USA.* **79**, 2608-2612.

Hama, H., Suzuki, K. & Tanaka, H. (1992); Inheritance and stability of resistance to *Bacillus thuringiensis* formulations of the diamondback moth, *Plutella xylostella* (*linaeus*) (Lepidoptera, Yponomeutidae). Appl. Entom. Zool. **27**, 355-362.

Hannay, C.L. (1953); Crystalline inclusions in aerobic spore-forming bacteria. *Nature* **172**, 1004-1005.

Harvey, W.R. & Wolferberger, M.G. (1979); Mechanism of inhibition of active potassium transport in isolated midgut of *Manduca sexta* by *Bacillus thuringiensis* endotoxin. *J. Exp. Biol.* **83**, 293-304.

Hawkes, R., Niday, E. & Gordon, J. (1982); A dot-immunobinding assay for monoclonal and other antibodies. *Anal. Biochem.* **119**, 142-147.

Heffron, F. (1983); Tn*3* and its relatives. In *Mobile Genetic Elements* (Shapiro, J.A., Ed), pp.223-261, Academic Press, New York, NY.

Heffron, F., Sublett, R., Hedges, R.W., Jacob, A & Falkow, W. (1975); Origin of the TEM Beta-lactamase gene found on plasmids; *J. Bacteriol.* **122**, 250-256.

Heffron, F., McCarthy, B.J., Ohtsubo, H. & Ohtsubu, E. (1979); DNA sequence analysis of the transposon Tn*3*; three genes and three sites involved in transposition of Tn*3*. *Cell* **18**, 1153-1164.

Heimpel, A.M. & Angus, T.A. (1959); The site of action of crystalliferous bacteria in lepidoptera larvae. *J. Invertebr. Pathol.* **1**, 152-170.

Heierson, A., Siden, I., Kivaisi, A. & Boman, H.G. (1986); Bacteriophage-resistant mutants of *Bacillus thuringiensis* with decreased virulence in pupae of *Hyalophora cecropia*. *J. Bacteriol.* **167**, 18-24.

Held, G.A., Kawanishi, C.Y. & Huang, Y.-S. (1990); Characterisation of the parasporal inclusion of *Bacillus thuringiensis* subsp. *israelensis*. *J. Bacteriol.* **172**, 481-483.

Held, G.A., Bulla, L.A., Ferrari, E., Hoch, J., Aronson, A.I. & Minnich, S.A. (1982); Cloning and localisation of the lepidopteran protoxin gene of *Bacillus thuringiensis* subsp. *kurstaki*. *Proc. Natl. Acad. Sci. USA.* **79**, 6065-6069.

Henikoff, S. (1984); Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351-359.

Henikoff, S. (1987); Unidirectional digestion with exonuclease III in DNA sequence analysis. *Methods Enzymol.* **155**, 156-165.

Herrnstadt, C., Soares, G.G., Wilcox, E.R. and Edwards, D.L. (1986); A new strain of *Bacillus thuringiensis* with activity against coleopteran insects. *Bio/Technology* **4**, 305-308.

Hodgman, T.C. & Ellar, D.J. (1990); Models for the structure and function of the *Bacillus thuringiensis* δ-endotoxins determined by complementation analysis. *Journal of DNA Mapping and Sequencing* **1**, 97-106.

Hodgman, T.C., Ziniu, Y., Shen, J. & Ellar, D.J. (1993); Identification of a cryptic gene associated with an Insertion Sequence not previously identified in *Bacillus thuringiensis*. *FEMS Microbiol. Lett.* **114**, 23-30.

Hofmann, C., Vanderbruggen, H., Höfte, H., Van Rie, J., Jansens, S. & Van Mellaert, H. (1988a); Specificity of *Bacillus thuringiensis* δ-endotoxins is correlated with the

presence of high-affinity binding sites in the brush border membrane of target insect midguts. *Proc. Natl. Acad. Sci. USA.* **85**, 7844-7848.

Hofmann, C., Lüthy, P., Hutter, R. and Pliska, V. (1988b); Binding of the delta-endotoxin from *Bacillus thuringiensis* to brush-border membrane vesicles of the cabbage butterfly (*Pieris brassicae*); Eur. J. Biochem. **173**, 85-91.

Höfte, H. & Whiteley, H.R. (1989); Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.* **53**, 242-255.

Höfte, H., de Grev, H., Seurinck, J., Jansens, S., Mahillon, J., Ampe, C., Vandekerckhove, J., Vanderbruggen, H., van Montagu, M., Zabeau, M. & Vaeck, M. (1986); Structural and functional analysis of a cloned delta-endotoxin of *Bacillus thuringiensis berliner* 1715. *Eur. J. Biochem.* **161**, 273-280.

Honée, G., van der Salm, T. & Visser, B. (1988); Nucleotide sequence of crystal protein gene isolated from *B. thuringiensis* subspecies *entomocidus* 60.5 coding for a toxin highly active against *Spodoptera* species. *Nucleic Acids Res.* **16**, 6240.

Huber, H.E., Lüthy, P., Ebersold, H.R. & Cordier, J.-L. (1981); The sub-units of the parasporal crystal of *Bacillus thuringiensis*: Size, linkage and toxicity. *Arch. Microbiol.* **129**, 14-18.

Ibarra, J.E. & Federici, B.A. (1986); Parasporal bodies of *Bacillus thuringiensis* subsp. *morrisoni* (PG-14) and *Bacillus thuringiensis* subsp. *israelensis* are similar in protein composition and toxicity. *FEMS Microbiol. Lett.* **34**, 79-84.

Iida, S., Meyer, J. & Arber, W. (1983); Prokaryotic IS elements; In *Mobile Genetic Elements* (Shapiro, J.A., Ed), pp.159-221, Academic Press, New York.

Illades-Aguiar, B. & Setlow, P. (1994); Autoprocessing of the protease that degrades small, acid-soluble protein of spores of *Bacillus* species is triggered by low pH, dehydration, and dipicolinic acid. *J. Bacteriol.* **176**, 7032-7037.

Ishii, T. & Ohba, M. (1993); Diversity of *Bacillus thuringiensis* environmental isolates showing larvicidal activity specific for mosquitoes. *J. Gen. Microbiol.* **139**, 2849-2854.

Jackson, S.G., Goodbrand, R.B., Ahmed, R. & Kasatiya, S. (1995); *Bacillus cereus* and *Bacillus thuringiensis* isolated in a gastroenteritis outbreak investigation. *Lett. Appl. Microbiol.* **21**, 103-105.

Jaquet, F., Hutter, R. & Lüthy, P. (1987); Specificity of *Bacillus thuringiensis* delta-endotoxin. *Appl. Environ. Microbiol.* **53**, 500-504.

Jenkin, H.M., McMeans, E., Anderson, L.E. & Yang, T.K. (1976); Phospholipid composition of *Culex quinquefasciatus* and *Culex tritaeniorhynchus* cells in logarithmic and stationary growth phases. *Lipids* **11**, 697-704.

Johnson, D.E. (1981); Toxicity of *Bacillus thuringiensis* entomocidal protein towards cultured insect tissue. *J. Invertebr. Pathol.* **38**, 94-101.

Johnson, D.E., Niezgodski, D.M. & Twaddle, G.M. (1980); Parasporal crystals produced by oligosporogenous mutants of *Bacillus thuringiensis* (Spo^-, Cr^+). *Can. J. Microbiol.* **26**, 486-491.

Johnson, W.C. & Tipper, D.J. (1981); Acid-soluble spore proteins of *Bacillus subtilis*. *J. Bacteriol.* **146**, 972-982.

Kaile-Gilliland, A., Bone, E. & Ellar, D.J. (1997); Microelectrode measurement of pH in the midgut of lepidopterans during larval development. In *30th Annual Meeting of the Society for Invertebrate Pathology*, Banff, Canada.

Kalman, S., Kiehne, K.L., Libs, J.L. & Yamamoto, T. (1993); Cloning of a novel *cryIC* type gene from a strain of *Bacillus thuringiensis* subsp. *galleriae*. *Appl. Environ. Microbiol.* **59**, 1131-1137.

Kalman, S., Kiehne, K.L., Cooper, N., Reynoso, M.S, & Yamamoto, T. (1995); Enhanced production of insecticidal proteins in *Bacillus thuringiensis* strains carrying an additional crystal protein gene in their chromosomes. *Appl. Environ. Microbiol.* **61**, 3063-3068.

Karamanlidou, G., Lambropoulos, A.F., Koliais, S.I., Manousis, T. & Ellar, D. (1991); Toxicity of *Bacillus thuringiensis* to laboratory populations of the olive fruit fly (*Dacus Oleae*). *Appl. Environ. Microbiol.* **57**, 2277-2282.

Kawalek, M.D., Benjamin, S., Lee, H.L. & Gill, S.S. (1995); Isolation and identification of novel toxins from a new mosquitocidal isolate from Malaysia, *Bacillus thuringiensis* ssp. *jegathesan*. *Appl. Environ. Microbiol.* **61**, 2965-2969.

Kirsch, K. & Schmutterer, H. (1988); Low efficacy of a *Bacillus thuringiensis* ssp. *berliner* formulation in controlling the Diamond back moth, *Plutella xylostella* in the Philippines. *J. Appl. Entomol.* Zeitschrift fur angewandte entomologie **105**, 249-255.

Klaer, R., Kuhn, S., Tillmann, E., Fritz, H.J. & Starlinger, P. (1981); The sequence of IS4. Mol. Gen. Genet. **181**, 169-175.

Kleckner, N. (1981); Transposable elements in prokaryotes. *Ann. Rev. Genet.* **191**, 341-404.

Klier, A., Fargette, F., Ribier, J. and Rapoport, G. (1982); Cloning and expression of the crystal protein genes from *Bacillus thuringiensis* strain *berliner* 1715. *EMBO J.* **1**, 791-799.

Knight, P.J.K., Crickmore, N. & Ellar, D.J. (1994); The receptor for *Bacillus thuringiensis* CryIA(c) delta-endotoxin in the brush border membrane of the lepidopteran *Manduca sexta* is aminopeptidase N. *Mol. Microbiol.* **11**, 429-436.

Knowles, B.H. (1994); Mechanism of action of *Bacillus thuringiensis* insecticidal δ-endotoxins. *Advances in Insect Physiology* **24**, 275-308.

Knowles, B.H., Blatt, M.R., Tester, M., Horsnell, J.M., Carroll, J.G., Menestrina, G. & Ellar, D.J. (1989); A cytolytic δ-endotoxin from *Bacillus thuringiensis* var. *israelensis* forms cation-selective channels in planar lipid bilayers. *FEBS Lett.* **244**, 259-262.

Knowles, B.H. & Ellar, D.J. (1986); Characterisation and partial purification of a plasma membrane receptor for *Bacillus thuringiensis* var. *kurstaki* lepidopteran-specific δ-endotoxin. *J. Cell Sci.* **83**, 89-101.

Knowles, B.H. & Ellar, D.J. (1987); Colloid-osmotic lysis is a general feature of the mechanism of action of *Bacillus thuringiensis* δ-endotoxins with different specificity. *Biochim. Biophys. Acta* **924**, 509-518.

Knowles, B.H., Knight, P.J.K. & Ellar, D.J. (1991); N-acetyl galactosamine is part of the receptor in insect gut epithelia that recognizes an insecticidal protein from *Bacillus thuringiensis*. *Proc. R. Soc. Lond. B* **245**, 31-35.

Knowles, B.H., White, P.J., Nicholls, C.N. & Ellar, D.J. (1992); A broad spectrum cytolytic toxin from *Bacillus thuringiensis* var. *kyushuensis*. *Proc. R. Soc. Lond.* **248**, 1-7.

Koni, P.A. & Ellar, D.J. (1993); Cloning and characterization of a novel *Bacillus thuringiensis* cytolytic delta-Endotoxin. *J. Mol. Biol.* **229**, 319-327.

Koni, P.A. & Ellar, D.J. (1994); Biochemical characterisation of *Bacillus thuringiensis* cytolytic delta-endotoxins. *Microbiol.* **140**, 1869-1880.

Kostichka, K., Warren, G.W., Mullins, M., Mullins, A.D., Craig, J.A., Kosiel, M.G. & Estruch, J.J. (1996); Cloning of a CryV-type insecticidal protein gene from *Bacillus thuringiensis*; the *cry*V encoded protein is expressed early in stationary phase. *J. Bacteriol.* **178**, 2141-2144.

Krieg, A. (1971); Concerning alpha-exotoxin produced by vegetative cells of *Bacillus thuringiensis* and *Bacillus cereus*. *J. Invertebr. Pathol.* **17**, 134-135.

Kröger, M. & Hobom, G. (1982); Structural analysis of insertion sequence IS*5*. *Nature* **297**, 159-162.

Kronstad, J.W. & Whiteley, H.R. (1984); Inverted repeat sequences flank a *Bacillus thuringiensis* crystal protein gene. *J. Bacteriol.* **160**, 95-102.

Kronstad, J.W., Schnepf, H.E. & Whiteley, H.R. (1983); Diversity of locations for *Bacillus thuringiensis* crystal protein genes. *J. Bacteriol.* **154**, 419-428.

Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P.G. & Skalka, A.M. (1992); Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.* **12**, 2231-2338.

Kunz, P.A. (1978); Resolution and properties of the proteinases in the larva of the mosquito, *Aedes aegypti*. *Insect Biochem.* **8**, 43-51.

Kvistgaard, M. (1994); Mikrobiologiske bekaempelsesmidler (In Danish). Miljoproject 256, 115 pp. Danish Environmental Protection Agency, Copenhagen, Denmark.

Lacey, L.A. & Federici, B.A. (1979); Pathogenesis and midgut histopathology of *Bacillus thuringiensis* in *Simulium vittatum* (Diptera: Simuliidae). *J. Invertebr. Pathol.* **33**, 171-182.

Lambert, B. & Peferoen, M. (1992); Insecticidal promise of *Bacillus thuringiensis*. Facts and mysteries about a successful biopesticide. *BioScience* **42**, 112-122.

Lampel, J.S., Canter, G.L., Dimock, M.B., Kelly, J.L,. Anderson, J.J., Uratani, B.B., Foulke, Jr., J.S. & Turner, J.T. (1994); Integrative cloning, expression and stability of the *cryIa(c)* gene from *Bacillus thuringiensis* subsp. *kurstaki* in a recombinant strain of *clavibacter-xyli* subsp. *cynodontis*. *Appl. Environ. Microbiol.* **60**, 501-508.

Lecadet, M.M. (1966); La toxine figurée de *Bacillus thuringiensis*: Dissolution par action du thioglycolate ou de la cysteine. *C. R. Acad. Sc. Paris Serie D* **262**, 195-198.

Lecadet, M.M. (1967); Action comparee de l'uree et du thioglycolate sur la toxine figuree de *Bacillus thuringiensis*. *C. R. Acad. Sc. Paris Serie D* **264**, 2847-2850.

Lecadet, M. M., Frachon, E., Dumanoir, V. C. & de Barjac, H. (1994); An updated version of the *Bacillus thuringiensis* strains classification according to H-serotypes. In *IInd International Conference on Bacillus thuringiensis*, Montpelier, France.

Lecadet, M.M. & Dedonder, R. (1971); Biogenesis of the crystalline inclusion of *Bacillus thuringiensis* during sporulation. *Eur. J. Biochem.* **23**, 282-294.

Lecadet, M.M. & Martouret, D. (1967); Enzymatic hydrolysis of the crystals of *Bacillus thuringiensis* by the proteases of *Pieris brassicae*: II. Toxicity of the different fractions of the hydrolysate for larvae of *Pieris brassicae*. *J. Invertebr. Pathol.* **9**, 322-330.

Lee, M. K., Milne, R. E., Ge, A. Z. & Dean, D. H. (1992); Location of a *Bombyx mori* receptor binding region on a *Bacillus thuringiensis* δ-endotoxin. *J. Biol. Chem.* **267**, 3115-3121.

Lee, M. J. & Anstee, J. H. (1995); Endoproteases from the midgut of larval *Spodoptera littoralis* include a chymotrypsin-like enzyme with an extended binding site. *Insect Biochem. Molec. Biol.* **25**, 49-61.

Leighton, T.J. & Stock, J.J. (1970). Biochemical changes during fungal sporulation and spore germination. *J. Bacteriol.* **101**, 931-940.

Léonard, C., Chen, Y. & Mahillon, J. (1997); Diversity and different distribution of IS*231*, IS*232* and IS*240* among *Bacillus cereus*, *Bacillus thuringiensis* and *Bacillus mycoides*. *Microbiology* **143**, 2537-2547.

Lereclus, D., Mahillon, J., Menou, G. & Lecadet, M.M. (1986); Identification of Tn*4430*, a transposon of *Bacillus thuringiensis* functional in *Escherichia coli*. *Mol. Gen. Genet.* **204**, 52-57.

Lereclus, D., Ribier, J., Klier, A., Menou, G. & Lecadet, M.M. (1984); A transposon-like structure related to the δ-endotoxin gene of *Bacillus thuringiensis*. *EMBO J.* **3**, 2561-2567.

Lereclus, D., Menou, G. & Lecadet, M-M. (1983); Isolation of a DNA sequence related to several plasmids from *Bacillus thuringiensis* after a mating involving *Streptococcus faecalis* pAMb1. *Mol. Gen. Genet.* **191**, 307-313.

Lereclus, D., Arantes, O., Chafaux, J. & Lecadet, M.M. (1989); Transformation and expression of a cloned δ-endotoxin gene in *Bacillus thuringiensis*. *Fems Microbiol. Lett.* **60**, 211-218.

Lereclus, D., Vallade, M., Chaufaux, J., Arantes, O. & Rambaud, S. (1992); Expansion of insecticidal host range of *Bacillus thuringiensis* by *in vivo* genetic recombination. *Bio/Technology* **10**, 418-421.

Levinson, B.L., Kasyan, K.J., Chiu, S.S., Currier, T.C. & Gonzalez, J.M. (1990); Identification of β-exotoxin, and a new exotoxin in *Bacillus thuringiensis* by using high-performance liquid chromatography. *J. Bacteriol.* **172**, 3172-3179.

Li, J., Carroll, J. & Ellar, D.J. (1991); Crystal structure of insecticidal δ-endotoxin from *Bacillus thuringiensis* at 2.5Å resolution. *Nature* **353**, 815-821.

Li, R.S., Jarrett, P. & Burges, H.D. (1987); Importance of spores, crystals, and δ-endotoxins in the pathogenicity of different varieties of *Bacillus thuringiensis* in *Galleria mellonella* and *Pieris brassicae*. *J. Invertebr. Pathol.* **50**, 277-284.

Li, J., Koni, P.A. & Ellar, D.J. (1996); Structure of the mosquitocidal δ-endotoxin CytB from *Bacillus thuringiensis* sp. *kyushuensis* and implications for pore membrane formation. *J. Mol. Biol.* **257**, 129-152.

Liang, Y., Patel, S.S. & Dean, D.H. (1995); Irreversible binding kinetics of *Bacillus thuringiensis* CryIA δ-endotoxins to gypsy moth brush border membrane vesicles is directly correlated to toxicity. *J. Biol. Chem.* **270**, 24719-24724.

Liu, Y.-T., Sui, M.-J., Ji, D.-D., Wu, I.-H., Chou, C.-C. & Chen, C.-C. (1993); Protection from ultraviolet light irradiation by melanin of mosquitocidal activity of *Bacillus thuringiensis* var. *israelensis*. *J. Invertebr. Pathol.* **62**, 131-136.

Lopez-Meza, J.E. & Ibarra, J.E. (1996); Characterization of a novel strain *of Bacillus thuringiensis*. *Appl. Environ. Microbiol.* **62**, 1306-1310.

Lorence, A., Darszon, A., Diaz, C., Lievano, A., Quintero, R. & Bravo, A. (1995); Delta-endotoxins induce cation channels in *Spodoptera frugiperda* brush-border membranes in suspension and in planar lipid bilayers. *FEBS Lett.* **360**, 217-222.

Loshon, C.A., Fliss, E.R., Setlow, B., Foerster, H.F. & Setlow, P. (1986); Cloning and nucleotide sequencing of genes for small, acid soluble spore proteins of *Bacillus cereus*, *Bacillus strearothermophilus* and *Thermoactinomyces thalophilus*. *J. Bacteriol.* **167**, 168-173.

Lövgren, A., Zhang, M.-Y, Engström, A., Dalhammar, G. & Landen, R. (1990); Molecular characterisation of immune inhibitor A, a secreted virulence protease from *Bacillus thuringiensis*. *Mol. Microbiol.* **4**, 2137-2146.

Lowry, O.H., Rosebrough, N.J., Farr, A.L. & Randall, R.J. (1951); Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**, 265-275.

Lu, H., Rajamohan, F. & Dean, D. H. (1994); Identification of amino acid residues of *Bacillus thuringiensis* δ-endotoxin CryIA(a) associated with membrane binding and toxicity to *Bombyx mori*. *J. Bacteriol.* **176**, 5554-5559.

Lukonen, A., Brummer-Korvenkontio, M. & Renkonen, O. (1973); Lipids of cultured mosquito cells (*Aedes albopictus*): comparison with cultured mammalian fibroblasts (BHK 21 cells). *Biochim. Biophys. Acta* **326**, 256-261.

Lüthy, P. (1980); Insecticidal toxins of *Bacillus thuringiensis*. *FEMS Microbiol. Lett.* **8**, 1-7.

Lüthy, P., Cordier, J.-L. & Fischer, H.-M. (1982); *Bacillus thuringiensis* as a bacterial insecticide: basic considerations and applications. In *Microbial and Viral Pesticides* (Kurstak, E., Ed.), pp. 35-74; Marcel Dekker, Inc., New York.

MacHattie, L.A. & Jackowski, J.B. (1977); Physical structure and deletion effects of the chloramphenicol resistant element Tn*9* in phage lambda. *In DNA Insertion Elements. Plasmids and Episomes* (Bukhari, A.I., Shapiro, J.A. and Adhya, S.L., Eds.), pp. 219-228, Cold Spring Harbor Press, New York.

Machesky, H. (1989); USDA forest service gypsy moth aerial suppression eradication projects. *Gypsy Moth News* **20**, 2-3.

Magill, N.G., Loshon, C.A. & Setlow, P. (1990); Small, acid-soluble, spore proteins and their genes from two species of *Sporosarcina*. *FEMS Microbiol. Lett.* **72**, 293-298.

Magill, N.G., Cowan, A.E., Koppel, D.E. & Setlow, P. (1994); The internal pH of the forespore compartment of *Bacillus megaterium* decreases by about 1 pH unit during sporulation. *J. Bacteriol.* **176**, 2252-2258.

Mahillon, J. & Lereclus, D. (1988); Structural and functional analysis of Tn*4430*: identification of an integrase-like protein involved in the co-integrate resolution process. *EMBO J.* **7**, 1515-1526.

Mahillon, J., Rezsohazy, R., Hallet, B. & Delcour, J. (1994); IS*231* and other *Bacillus thuringiensis* transposable elements: a review. *Genetica* **93**, 13-26.

Mahillon, J., Seurinck, J., Van Rompuy, L., Delcour, J. & Zabeau, M. (1985); Nucleotide sequence and structural organisation of an insertion sequence element (IS*231*) from *Bacillus thuringiensis* strain *berliner* 1715. *EMBO J.* **4**, 3895-3899.

Mahillon, J., Seurinck, J., Delcour, J. & Zabeau, M. (1987); Cloning and nucleotide sequence of different iso-IS*231* elements and their structural association with the Tn*4430* transposon in *Bacillus thuringiensis*. *Gene* **51**, 187-196.

Mahillon, J., Chungjatupornchai, W., Decock, J., Dierickx, S., Michiels, F., Peferoen, M. & Joos, H. (1989); Transformation of *Bacillus thuringiensis* by electroporation. *FEMS Microbiol. Lett.* **60**, 205-210.

Manasherob, R., Bendov, E., Zaritsky, A. & Bartak, Z. (1994); Protozoan enhanced toxicity of *Bacillus thuringiensis* var. *israelensis* δ-endotoxin against *Aedes aegypti* larvae. *J. Invertebr. Pathol.* **63**, 244-248.

Manasherob, R., BenDov, E., Margalit, J., Zaritsky, A. & Barak, Z. (1996); Raising activity of *Bacillus thuringiensis* var. *israelensis* against *Anopheles stephensi* larvae by encapsulation in *Tetrahymena pyriformis* (Hymenostomatida: Tetrahymenidae). *J. Am. Mosq. Cont. Assoc.* **12**, 627-631.

Mason, J.M. & Setlow, P. (1986); Different small, acid soluble proteins of the α/β-type have interchangeable roles in the heat and UV radiation resistance of *Bacillus subtilis* spores. *J. Bacteriol.* **169**, 3633-3637.

Mason, J.M. & Setlow, P. (1987); Essential role of small, acid-soluble spore proteins in resistance of *Bacillus subtilis* to UV light. *J. Bacteriol.* **167**, 174-178.

McGaughey, W.H. (1985); Insect resistance to the biological insecticide *Bacillus thuringiensis*. *Science* **229**, 193-194.

McGaughey, W.H. & Johnson, D.E. (1994); Influence of crystal protein composition of *Bacillus thuringiensis* strains on cross-resistance in Indian meal moths (Lepidoptera, pyralidae). *J. Econ. Entomol.* **87**, 535-540.

McLean, K.M. & Whiteley, H.R. (1987); Expression in *Escherichia coli* of a cloned crystal protein gene of *Bacillus thuringiensis* subsp. *israelensis*. *J. Bacteriol.* **169**, 1017-1023.

Meadows, M.P., Ellis, D.J., Butt, J., Jarret, P. & Burges, H.D. (1992); Distribution, frequency and diversity of *Bacillus thuringiensis* in an animal feed mill. *Appl. Environ. Microbiol.* **58**, 1344-1350.

Menou, G., Mahillon, J., Lecadet, M.-M. & Lereclus, D. (1990); Structural and genetic organisation of IS*232*, a new insertion sequence of *Bacillus thuringiensis*. *J. Bacteriol.* **172**, 6689-6696.

Miyasono, M., Inagaki, S., Yamamoto, M., Ohba, K., Ishiguro, T., Takeda, R. and Hayashi, Y. (1994); Enhancement of δ-endotoxin activity by toxin-free spore of *Bacillus thuringiensis* against the diamondback moth, *Plutella xylostella*. *J. Invertebr. Pathol.* **63**, 111-112.

Moar, W.J., Masson, L., Brousseau, R. & Trumble, J.T. (1990); Toxicity to *Spodoptera exigua* and *trichoplusia ni* of individual P1 protoxins and sporulated cultures of *Bacillus thuringiensis* ssp. *kurstaki* HD-1 and NDR-12. *Appl. Environ. Microbiol.* **56**, 2480-2483.

Mollet, B., Iida, S., Sepherd, J. & Arber, W. (1983); Nucleotide sequence of IS*26*, a new prokaryotic mobile genetic element. *Nucl. Acids Res.* **11**, 6319-6330.

Moran, C.P. (1993); RNA polymerase and transcription factors. In *Bacillus subtilis and other Gram Positive Bacteria* (Sonenshein, A.L., Hoch, J.A. and Losick, R. Eds.), pp. 653-667, American Society for Microbiology, Washington DC.

Moran, C.P., Jr., Lang, N. & Losick, R. (1981); Nucleotide sequence of a *Bacillus subtilis* promoter recognized by *Bacillus subtilis* of a *Bacillus subtilis* RNA polymerase containing σ^{37}. *Nucleic Acids Res.* **9**, 5979-5990.

Monroy, A, Maggio, R. & Rinaldi, A.M. (1965); Experimentally induced activation of the ribosomes of the unfertilized sea urchin egg. *Proc. Natl. Acad. Sci. USA.* **54**, 107-111.

Munoz, L., Sadaie, Y., Sadaie, Y. & Doi, R. (1978); *J. Biol. Chem.* **253**, 6994-6701.

Murphy, D. W., Sohi, S. S. & Fast, P. G. (1976); *Bacillus thuringiensis* enzyme-digested delta-endotoxin: Effect on cultured insect cells. *Science* **194**, 954-956.

Nagamatsu, Y., Itai, Y., Hatanaka, C., Funatsu, G. & Hayashi, K. (1984); A toxic fragment from the entomocidal crystal protein of *Bacillus thuringiensis*. *Agric. Biol. Chem.* **48**, 611-619.

Nakamura, I.K. (1994); DNA relatedness among *Bacillus thuringiensis* serovars. *Int. J. Syst. Bacteriol.* **44**, 125-129.

Nakayama, T., Munoz, L., Sadaie, Y. & Doi, R. (1978); Spore coat protein synthesis in cell-free systems from sporulating cells of *Bacillus subtilis*. *J. Bacteriol.* **135**, 952-960.

Nicholls, C.N., Ahmad, W. & Ellar, D.J. (1989); Evidence for two different types of insecticidal P2 toxins with dual specificity in *Bacillus thuringiensis* subspecies. *J. Bacteriol.* **171**, 5141-5147.

Nicolas, L., Charles, J.-F. & de Barjac, H. (1993); *Clostridium bifermentans* serovar *malaysia*: characterisation of putative mosquito larvicidal proteins. *FEMS Microbiol. Lett.* **113**, 23-28.

Nicholson, W.L., Setlow, B. & Setlow, P. (1990); Binding of DNA *in vitro* by a small, acid soluble spore protein from *Bacillus subtilis* and the effect of this binding on DNA topology. *J. Bacteriol.* **172**, 6900-6906.

Nishiitsutsuji-Uwo, J., Endo, Y. & Himeno, M. (1980); Effects of *Bacillus thuringiensis* δ-endotoxin on insect and mammalian cells *in vitro*. *Appl. Entomol. Zool.* **15**, 133-139.

Nishimoto, T., Yoshisue, H., Ihara, K., Sakai, H. & Komano, T. (1994); Functional analysis of block-5, one of the highly conserved amino-acid sequences in the 130-kDa CryIVA protein produced by *Bacillus thuringiensis* ssp. *israelensis*. *FEBS Lett.* **348**, 249-254.

Ohba, M. & Aizawa, K. (1979); Distribution of *Bacillus thuringiensis* serotypes in Ehime Prefecture, Japan. *Appl. Entomol. Zool.* **14**, 340-345.

Ohba, M. & Aizawa, K. (1989); New flagellar (H) antigenic subfactors in *Bacillus thuringiensis* H serotype-3 with description of 2 new subspecies, *Bacillus thuringiensis* subsp. *sumiyoshiensis* (H-serotype 3A-3D) and *Bacillus thuringiensis* subsp. *fukuokaensis* (H-serotype 3A-3D-3E). *J. Invertebr. Pathol.* **54**, 208-212.

Ohba, M. (1996); *Bacillus thuringiensis* populations naturally occurring on mulberry leaves: A possible source of the populations associated with silkworm rearing insectaries. *J. Appl. Bacteriol.* **80**, 56-64.

Orduz, S., Roja, W., Correa, M.M., Montoya, A.E. & de Barjac, H. (1992); A new serotype of *Bacillus thuringiensis* from Colombia toxic to mosquito larvae. *J. Invertebr. Pathol.* **59**, 293-294.

Padua, L.E., Ohba, M. & Aizawa, K. (1980); The isolates of *Bacillus thuringiensis* serotype 10 with a highly preferential toxicity to mosquito larvae. *J. Invertebr. Pathol.* **36**, 180-186.

Padua, L.E., Ohba, M. & Aizawa, K. (1984); Isolation of a *Bacillus thuringiensis* strain (serotype 8a:8b) highly and selectively toxic against mosquito larvae. *J. Invertebr. Pathol.* **44**, 12-17.

Pederson, J.C., Damgaard, P.H., Eilenberg, J. & Hansen, B.M. (1995); Dispersion of *Bacillus thuringiensis* var *kurstaki* in an experimental cabbage field. *Can J. Microbiol.* **41**, 118-125.

Peferoen, M. (1997); Progress and prospects for field use of Bt genes in crops. *Trends Biotechnol.* **15**, 173-177.

Pendleton, I.R., Bernheimer, A.W. & Grushoff, P. (1973); Purification and characterization of haemolysins from *Bacillus thuringiensis*. *J. Invertebr. Pathol.* **21**, 131-135.

Peters, W. (1992); Peritrophic membranes. *Zoophyiology* **30**, 119-122.

Petit-Glatron, M.F. & Rapoport, G. (1976); Translation of a stable mRNA fraction from sporulating cells of *Bacillus thuringiensis* in a cell-free system from *Escherichia coli*. *Biochimie* **58**, 119-129.

Polzin, K.M. & Shimizu-Kadota, M. (1987); Identification of a new insertion element, similar to Gram-negative IS*26*, on the lactose plasmid of *Streptococcus lactis* ML3. *J. Bacteriol.* **169**, 5481-5488.

Poncet, S. (1993); Etude et amelioration des proprietes biopesticides de bacteries entomopathogenes, utilisees comme moyen de lutte contre les diptres vecteurs des maladies tropicales. PhD Thesis, Université Paris, France.

Postemsky, C.J., Dignam, S.S. & Setlow, P. (1978); Isolation and characterisation of mutants of *Bacillus megaterium* containing decreased levels of spore protease. *J. Bacteriol.* **135**, 841-850.

Pujol, C., Ehrlich, S.D. & Janniere, L. (1994); The promiscuous plasmids pIP501 and pAMb1 from Gram-positive bacteria encode complementary resolution functions. *Plasmid* **31**, 100-105.

Purcell, J.P., Greenplate, J.T. & Sammons, R.D. (1992); Examination of midgut luminal proteinase activities in six economically important insects. *Insect Biochem. Molec. Biol.* **22**, 41-47.

Purcell, J.P., Greenplate, J.T., Jennings, M.G., Ryerse, J.S., Pershing, J.C., Sims, S.R., Prinsen, M.J., Corbin, D.R., Tran, M., Sammons, R.D. & Stonard, R.J. (1993); Cholesterol oxidase - a potent insecticidal protein active against boll weevil larvae. *Biochemical and Biophysical Research Communications* **196**, 1406-1413.

Purcell, M.D. (1997); The mosquitocidal activity of *Bacillus thuringiensis* ssp. *israelensis*. PhD Thesis, University of Cambridge.

Pusztai, M., Fast, P., Gringortin, L., Kaplan, H., Lessard, T. & Carey, P.R. (1991); The mechanism of sunlight-mediated inactivation of *Bacillus thuringiensis* crystals. *Biochem. J.* **273**, 43-47.

Ragni, A., Thiery, I. & Delécluse, A. (1996); Characterisation of six highly mosquitocidal *Bacillus thuringiensis* strains that do not belong to H-serotype. *Curr. Microbiol.* **32**, 48-54.

Rahardja, U. & Whalon, M.E. (1995); Inheritance of resistance *to Bacillus thuringiensis* ssp. *tenebrionis* CryIIIA δ-endotoxin in Colorado potato beetle. *J. Econ. Entomol.* **88**, 21-26.

Rajalkashmi, S. & Shethna, Y.I. (1977); The effect of amino acids on growth, sporulation and crystal formation in *Bacillus thuringiensis* var. *thuringiensis*. *J. Indian Inst. Sci.* **59**, 16.

Rang, C., Bes, M., Lullien-Pellerin, V., Wu, D., Federici, B.A. & Frutos, R. (1996); Influence of the 20-kDa protein from *Bacillus thuringiensis* ssp. *israelensis* on the rate of production of truncated Cry1C proteins. *FEMS Microbiol. Lett.* **141**, 261-264.

Regev, A., Keller, M., Strizhov, N., Sneh, B., Prudovsky, E., Chet, I., Ginzberg, I., Koncz-Kalman, Z., Koncz, C., Schell, J. & Zilberstein, A. (1996); Synergistic activity of a *Bacillus thuringiensis* delta-endotoxin and a bacterial endochitinase against *Spodoptera littoralis* larvae. *Appl. Environ. Microbiol.* **62**, 3581-3856.

Restrepo, N., Gutierrez, D., Patino, M.M., Thiery, I., Delécluse, A. & Orduz, S. (1997); Cloning, expression and toxicity of a mosquitocidal toxin gene *of Bacillus thuringiensis* subsp *medellin*. *Memorias do Instituto Oswaldo Cruz* **92**, 257-262.

Ribier, J. & Lecadet, M.M. (1973); Etude ultrastructurale et cinetique de la sporulation de *Bacillus thuringiensis* var. *berliner* 1715. Remarques sur la formation de l'inclusion parasporale. *Ann. Microbiol.* (Inst. Pasteur) **124A**, 311-344.

Rosso, M.-L. & Delécluse, A. (1997a); Contribution of the 65-kilodalton protein encoded by the cloned gene cry19A to the mosquitocidal activity of Bacillus thuringiensis subsp. jegathesan. *Appl. Environ. Microbiol.* **63**, 4449-4455.

Rosso, M.-L. & Delécluse, A. (1997b); Distribution of the insertion element IS*240* among *Bacillus thuringiensis* strains. *Curr. Microbiol.* **34**, 348-353.

Rowe, G.E. & Margaritis, A. (1987); Bioprocess developments in the production of bio-insecticides by *Bacillus thuringiensis*. *CRC Crit. Rev. Biotech.* **6**, 87-127.

Ryan, M., Johnson, J.D. & Bulla Jr., L.A. (1993); Insertion sequence elements in *Bacillus thuringiensis* subsp. *darmstadiensis*. *Can. J. Microbiol.* **39**, 649-658.

Ryter, A. (1965); Etude Morphologique de la sporulation de *Bacillus subtilis*. *Annales de l'Institut Pasteur* **108**, 40-60.

Sacchi, V.F., Parenti, P., Hanozet, G.M., Giordana, B., Luthy, P. & Wolfersberger, M (1986); *Bacillus thuringiensis* toxin inhibits K^+ gradient-dependent amino acids transport across the brush border membrane of *Pieris brassicae* midgut cells. *FEBS Lett.* **204**, 213-218.

Sakakibara, Y., Saito, H. Ikeda, Y. (1965). Incorporation of radioactive amino acids and bases into nucleic acid and protein fractions of germinating spores of *Bacillus subtilis*. *J. Gen. Appl. Microbiol.* **11**, 243-254.

Salama, H.S., Zaki, F.N., Sharaby, A. (1982); Effects of *Bacillus thuringiensis* ssp. *berliner* on parasites and predators of the cotton leafworm *Spodoptera littoralis*. *Z. ang. Ent.* **94**, 498-509.

Sanchez-Salas, J.L. & Setlow, P. (1993); Proteolytic processing of the protease which initiates degradation of small, acid-soluble proteins during germination of *Bacillus subtilis* spores. *J. Bacteriol.* **175**, 2568-2577.

Sanchis, V., Lereclus, D., Menou, G., Chaufaux, J. & Lecadet, M.M. (1988); Multiplicity of δ-endotoxin genes with different insecticidal specificities in *Bacillus thuringiensis aizawai* 7.29. *Mol. Microbiol.* **2**, 393-404.

Sanchis, V., Lereclus, D., Menou, G., Chaufaux, J., Guo, S. & Lecadet, M.M. (1989); Nucleotide sequence and analysis of the N-terminal coding region of the *Spodoptera*-active δ-endotoxin gene of *Bacillus thuringiensis aizawai* 7.29. *Mol. Microbiol.* **3**, 229-238.

Sangadala, S., Walters, F.S., English, L.D. & Adang, M.J. (1994); A mixture of *Manduca sexta* aminopeptidase and phosphatase enhances *Bacillus thuringiensis* insecticidal CryIA(c) toxin binding and RB-86 efflux *in vitro*. *J. Biological. Chem.* **269**, 10088-10092.

Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. & Vijayan, M. (1996); A novel mode of carbohydrate recognition in jacalin, a *Moraceae* plant lectin with a β-prism fold. *Nature Struct. Biol.* **3**, 596-603.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989); *Molecular Cloning: A Laboratory Manual*. 2nd edit, Cold Spring Harbor Laboratory Press, New York.

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zwigler, D.R. & Dean, D.H. (1998); *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol. Mol. Biol. Rev.* **62**, 775-806.

Schnepf, H.E. & Whiteley, H.R. (1981); Cloning and expression of the *Bacillus thuringiensis* crystal protein gene in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*. **78**, 2893-2897.

Schnepf, H.E. & Whiteley, H.R. (1985); Delineation of a toxin-encoding segment of a *Bacillus thuringiensis* crystal protein gene. *J. Biol. Chem.* **260**, 6273-6280.

Schnepf, H.E., Tomczak, K., Ortega, J.P. & Whiteley, H.P. (1990); Specificity determining regions of a lepidopteran-specific insecticidal protein produced by *Bacillus thuringiensis*. *J. Biol. Chem.* **265**, 20923-20930.

Schwartz, J.-L., Garneau, L., Savaria, D., Masson, L., Brousseau, R. & Rousseau, E. (1993); Lepidopteran-specific crystal toxins from *Bacillus thuringiensis* form cation- and anion-selective channels in planar lipid bilayers. *J. Membr. Biol.* **132**, 53-62.

Schwartz, E., Kroger, M. & Rak, B. (1988); IS*150*: distribution, nucleotide sequence and phylogenetic relationships of a new *E. coli* insertion element. *Nucl. Acids Res.* **16**, 6789-6802.

Sêbesta, K., Farkas, J., Horska, K. & Vankova, J. (1981); Thuringiensin, the beta-exotoxin of *Bacillus thuringiensis*. In *Microbial Control of Pests and Plant Diseases 1970-1980* (Burges, H. D., Ed.), pp. 249-281, Academic Press, Inc., London.

Sêbesta, K. & Sternbach, H. (1970); The specificity of inhibition of DNA-dependent RNA polymerase by *Bacillus thuringiensis* exotoxin. *FEBS Lett.* **8**, 233-235.

Sekar, V. (1988); The insecticidal crystal protein gene is expressed in vegetative cells of *Bacillus thuringiensis* var. *tenebrionis*. *Curr. Microbiol.* **17**, 347-349.

Sekar, V., Thompson, D.V., Maroney, M.J., Bookland, R.G. & Adang, M.J. (1987); Molecular cloning and characterization of the insecticidal crystal protein gene of *Bacillus thuringiensis* var. *tenebrionis*. *Proc. Natl. Acad. Sci. USA.* **84**, 7036-7040.

Setlow, P. (1975a); Purification and properties of some unique low molecular weight basic proteins degraded during germination of *Bacillus megaterium* spores. *J. Biol. Chem.* **250**, 8168-8173.

Setlow, P. (1975b); Protein metabolism during germination of *Bacillus megaterium* spores; degradation of pre-existing and newly synthesized protein. *J. Biol. Chem.* **250**, 631-637.

Setlow, P. (1975c); Identification and localization of the major proteins degraded during germination of *Bacillus megaterium* spores. *J. Biol. Chem.* **250**, 8159-8167.

Setlow, P. (1978); Purification of characterisation of additional low-molecular weight basic proteins degraded during germination of *Bacillus megaterium* spores. *J. Bacteriol.* **136**, 331-340.

Setlow, P. (1985); Protein degradation during bacterial spore germination. In *Fundamental and Applied Aspects of Bacterial Spores*, ed. D.J. Ellar, pp285-96. London: Academic.

Setlow, P. (1988); Small, acid-soluble proteins of *Bacillus* species: structure, synthesis, genetics, function and degradation. *Ann. Rev. Microbiol.* **42**, 319-338.

Setlow, P. (1992); I will survive - protecting and repairing spore DNA. *J. Bacteriol.* **174**, 2737-2741.

Setlow, P. & Kornberg, A. (1969); Biochemical studies of bacterial sporulation and germination. XVII. Sulfhydryl and disulfide levels in dormancy and germination. *J. Bacteriol.* **100**, 1155-1160.

Setlow, P. & Kornberg, A. (1970); Biochemical studies of bacterial sporulation and germination. XXII. Energy metabolism in early stages of germination of *Bacillus megaterium*. *J. Biol. Chem.* **245**, 2645-3652.

Setlow, P. & Ozols, J. (1979); Covalent structure of protein A: a low molecular weight protein degraded during germination of *Bacillus megaterium* spores. *J. Biol. Chem.* **254**, 11938-11942.

Setlow, P. & Ozols, J. (1980); Covalent structure of Protein C. A second major low molecular weight protein degraded during germination of *Bacillus megaterium* spores. *J. Biol. Chem.* **255**, 8413-8416.

Setlow, P. & Primus, G. (1975); Protein metabolism during germination of *Bacillus megaterium* spores. *J Biol. Chem.* **250**, 623-630.

Setlow, P. & Waites, W. (1976); Identification of several unique, low-molecular weight basic proteins in dormant spores of *Clostridium bifermentans* and their degradation during spore germination. *J. Bacteriol.* **127**, 1015-1017.

Setlow, P. Gerard, C. & Ozols, J. (1980); The amino acid sequence specificity of a protease from spores of *Bacillus megaterium. J. Biol. Chem.* **255**, 3624-3628.

Shapiro, J.A. (1979); Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proc. Natl. Acad. Sci. USA* **76**, 1933-1937.

Shelton, A.M., Robertson, J.L., Tang, J.D., Perez, C., Eigenbrode, S.D., Preisler, H.K., Wilsey, W.T., & Cooley, R.J. (1993); Resistance of Diamondback moth (Lepidoptera, plutellidae) to *Bacillus thuringiensis* subspecies in the field. *J. Econom. Entomol.* **86**, 697-705.

Shimuzu, T., Vassylyev, D.G., Kido, S. & Morkawa, K. (1994); Crystal structure of vitelline membrane outer layer protein I (VMO-I); a folding motif with homologous Greek key structures related by an internal three-fold symmetry. *EMBO J.* **15**, 1003-1010.

Shine, J. & Dalgarno, L. (1974); The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA.* **71**, 1342-1346.

Singh, R.P., Setlow, B. & Setlow, P. (1977); Levels of small molecules and enzymes in the mother cell compartment and the forespore of sporulating *Bacillus megaterium. J. Bacteriol.* **130**, 1130-1138.

Skøt, L,. Harrison, S.P., Nath, A., Mytton, L.R. & Clifford, B.C. (1990); Expression of insecticidal activity in rhizobium containing the δ-endotoxin gene cloned from *Bacillus thuringiensis* ssp. *tenebrionis. Plant and Soil* **127**, 285-295.

Slatin, S.L., Abrams, C.K. & English, L. (1990); Delta-endotoxins form cation-selective channels in planar lipid bilayers. *Biochim. Biophys. Res. Comm.* **169**, 765-772.

Smedley, D.P., Armstrong, G. & Ellar, D.J. (1997); Channel activity caused by *a Bacillus thuringiensis* δ-endotoxin preparation depends on the method of activation. *Molecular Membrane Biology* **14**, 13-18.

Smith, G.P. (1992); A study of the structure-function relationships of *Bacillus thuringiensis* δ-endotoxin. PhD thesis, University of Cambridge.

Smith, G.P. & Ellar, D.J. (1994); Mutagenesis of two surface-exposed loops of the *Bacillus thuringiensis* CryIC δ-endotoxin affects the insecticidal specificity. *Biochem. J.* **302**, 611-616.

Smith, G.P., Ellar, D.J., Keeler, S. & Seip, C.E. (1994); Nucleotide sequence and analysis of an insertion sequence from *Bacillus thuringiensis* related to IS*150*. *Plasmid* **32**, 10-18.

Smith, G.P., Merrick, J.D., Bone, E.J. & Ellar, D.J. (1996); Mosquitocidal activity of the CryIC δ-endotoxin from *Bacillus thuringiensis* subsp. *aizawai*. *Appl. Environ. Microbiol.* **62**, 680-684.

So, M., Heffron, F. & McCarthy, B.J. (1979); The *E. coli* gene encoding heat stable toxin is a bacterial transposon flanked by inverted repeats of IS*1*. *Nature* **277**, 453-456.

Soltesrak, E., Kushner, D.J., Williams, D.D. & Coleman, J.R. (1995); Factors regulating CryIVB expression in the Cyanobacterium *Synchococcus* PCC-7942. Mol. Gen. Genet. **246**, 301-308.

Staden, R. (1984); A computer program to enter DNA gel reading data into a computor. *Nuc. Acid. Res.* **12**, 499-503.

Staden, R. (1987); Computer handling of DNA sequencing projects. In *Nucleic acids and Protein Sequence Analysis: a Practical Approach* (Bishop, M.J. and Rawlings, C.J., Eds.), pp. 173-218; IRL Press, Oxford.

Stragier, P. & Losick, R. (1990); Cascades of sigma factors revisited. *Mol. Microbiol.* **2**, 1801-1806.

Stahly, D. P., Dingman, D. W., Bulla, L. A. & Aronson, A. I. (1978); Possible origin and function of the parasporal crystals in *Bacillus thuringiensis*. *Biochem. Biophys. Res. Commun.* **84**, 581-588.

Stark, W.M., Boocock, M.R. & Sherratt, D.J. (1992); Catalysis by site-specific recombinases. *Trends Genet.* **8**, 432-439.

Steffan, R.J., Breen, A., Atlas, R.M., & Sayler, G.S. (1989); Application of gene probe methods for monitoring specific microbial populations in fresh water ecosystems. *Can. J. Microbiol.* **35**, 681-685.

Stephens, J.M. (1952); Disease in codling moth larvae produced by several strains of *Bacillus cereus*. *Can. J. Zool.* **30**, 30-40.

Stewart, G.S.A.B,. Johnstone, K., Hagelberg, E. & Ellar, D.J.. (1981); Commitment of bacterial spores to germinate. *Biochem. J.* **198**, 101-106.

Stewart, C.N., All, J.N.Jr., Raymer, P.L. & Ramachandran, S. (1997); Increased fitness of transgenic insecticidal rapeseed under insect selection pressure. *Mol. Ecol.* **6**, 773-779.

Stone, T.B., Sims, S.R. & Marrone, P.G. (1989); Selection of tobacco budworm to a genetically engineered *Pseudomonas fluorescens* containing the δ-endotoxin of *Bacillus thuringiensis* subsp. *kurstaki*. *J. Invert. Pathol.* **53**, 228-234.

Sun, D.X. & Setlow, P. (1987); Cloning and nucleotide sequence of genes for a second type of small, acid-soluble spore proteins of *Bacillus thuringiensis*, *Bacillus strearothermophilus*, and *Thermoactinomyces thalophilus*. *J. Bacteriol.* **169**, 3088-3093.

Sussman, M.D. & Setlow, P. (1991); Cloning, nucleotide sequence, and regulation of the *Bacillus subtilis* GPR gene, which codes for the protease that initiates degradation of small, acid-soluble protein during spore germination. *J. Bacteriol.* **173**, 291-300.

Swinfield, T.-J., Janniere, L., Ehrlich, S.D. & Minton, N.P. (1991); Characterisation of a region of the *Enterococcus faecalis* plasmid pAMb1 which enhances the segregational stability of pAMb1-derived cloning vectors in *Bacillus subtilis*. *Plasmid* **26**, 209-221.

Tabashnik, B.E., Cushing, N.L., Finson, N & Johnson, M.W. (1990); Field development of resistance to *Bacillus thuringiensis* in diamondback moth (*Lepidoptera plutellidae*). *J Econ. Entomol.* **83**, 1671-1676.

Tabashnik, B.E., Finson, N. & Johnson, M.W. (1991); Managing resistance to *Bacillus thuringiensis* - lessons from the Diamondback moth (Lepidoptera, Plutellidae). *J. Econ. Entom.* **84**, 49-55.

Tabashnik, B.E., Finson, N., Johnson, M.W. & Moar, W.J. (1993); Resistance to toxins from *Bacillus thuringiensis* subsp. *kurstaki* causes minimal cross resistance to *Bacillus thuringiensis* subsp. *aizawai* in the Diamondback Moth (Lepidoptera: Plutellidae). *Appl. Environ. Microbiol.* **59**, 1332-1335.

Tabashnik, B.E., Malvar, T., Liu, Y.B., Finson, N., Borthakur, D., Shin, B.S., Park, S.H., Masson, L., DeMaagd, R.A. & Bosch, D. (1996); Cross-resistance of the diamondback moth indicates altered interactions with domain II of *Bacillus thuringiensis* toxins. *Appl. Environ. Microbiol.* **62**, 2839-2844.

Taguchi, R., Asahi, Y. & Ikezawa, H. (1980); Purification and properties of phosphatidylinositol specific phopholipase C of *Bacillus thuringiensis*. *Biochim. Biophys. Acta* **619**, 48-57.

Takesue, S., Yokota, K., Miyajima, S., Taguchi, R., Ikezawa, H. & Takesue, Y. (1992); Partial release of aminopeptidase N from larval midgut cell membranes of the silkworm, *Bombyx mori*, by phosphatidylinositol-specific *phospholipase C*. *Comp. Biochem. Physiol.* **102B**, 7-11.

Tapp, H. & Stotzky, G. (1995); Insecticidal activity of the toxins from *Bacillus thuringiensis* subspecies *kurstaki* and *tenebrionis* adsorbed and bound on pure and soil clays. *Appl. Environ. Microbiol.* **61**, 1786-1790.

te Giffel, M.C., Beumer, R.R., Klijn, N., Wagendorp, A. & Rombouts, F.M. (1997); Discrimination between *Bacillus cereus* and *Bacillus thuringiensis* using specific DNA probes based on variable regions of 16S rRNA. *FEMS Microbiol. Lett.* **146**, 47-51.

Teixeira de Souza, M., Lecadet, M.-M & Lereclus, D. (1993); Full expression of the *cryIIIA* toxin gene of *Bacillus thuringiensis* requires a distant upstream DNA sequence affecting transcription. *J. Bacteriol.* **175**, 2952-2960.

Thanabalu, Y., Hindley, J., Brenner, S., Oei, C. & Berry, C. (1992); Expression of the mosquitocidal toxins of *Bacillus sphaericus* and *Bacillus thuringiensis* subsp. *israelensis* by recombinant *Caulobacter crescentus*, a vehicle for biological control of aquatic insect larvae. *Appl. Environ. Microbiol.* **58**, 905-910.

Thie, N.M.R. & Houseman, J.G. (1990); Identification of cathepsin B, D and H in the larval midgut of Colorado Potato Beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Insect Biochem.* **20**, 313-318.

Thomas, W.E. & Ellar, D.J. (1983); *Bacillus thuringiensis* var. *israelensis* crystal δ-endotoxin: Effects on insect and mammalian cells *in vitro* and *in vivo*. *J. Cell Sci.* **60**, 181-197.

Thorne, L., Garduno, F., Thompson, T., Decker, D., Zounes, M., Wild, M., Walfield, A.M. & Pollock, T. (1986); Structural similarity between the Lepidoptera- and Diptera-specific insecticidal endotoxin genes from *Bacillus thuringiensis* subsp. *kurstaki* and *israelensis*. *J. Bacteriol.* **166**, 801-811.

Timmerman, K.P. & Tu, C.-P. (1985); Complete sequence of IS*3*. *Nucl. Acids Res.* **13**, 2127-2139.

Tomasino, S.F., Leister, R.T,. Dimock, M.B., Beach, R.M. & Kelly, J.L. (1995); Field performance of *Clavibacter xyli* ssp. *cynodontis* expressing the insecticidal crystal protein gene *cry1Ac* of *Bacillus thuringiensis* against European corn borer in field corn. *Biol. Controls* **5**, 442-448.

Tovar-Rojo, F. & Setlow, P. (1991); Effects of mutant small, acid-soluble spore proteins from *Bacillus subtilis* on DNA *in vivo* and *in vitro*. *J. Bacteriol.* **173**, 4827-4835.

Towbin, H., Staehelin, T. & Gordon, J. (1979); Electrophoretic transfer of proteins from acrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci.* **76**: 4350-4354.

Trieu-Cuot, P. & Courvalin, P. (1984); Nucleotide sequence of the transposable element IS*15*. *Gene* **30**, 113-120.

Vadlamudi, R.K., Weber, E., Ji, I., Ji, T.H. & Bulla, L.A., Jr. (1995); Cloning and expression of a receptor for an insecticidal toxin of *Bacillus thuringiensis*. *J. Biol. Chem.* **270**, 5490-5494.

Vaeck, M., Reynaerts, A., Hofte, H., Jansens, S., De Beuckeleer, M., Dean, C., Zabeau, M., Van Montagu, M & Leemans (1987); Transgenic plants protected from insect attack. *Nature* **328**, 33-37

Valetis, A.P., Mazza, A., Brousseau, R. & Masson, L. (1995); Interaction analyses of *Bacillus thuringiensis* Cry1A toxins with two aminopeptidases from gypsy moth midgut brush border membranes. *Insect Biochem. Mol. Biol.* **27**, 529-539

Van Frankenhuyzen, K., Gringorten, J.L., Gauthier, D., Milne, R.E. & Masson, L. (1993); Toxicity of activated CryI proteins from *Bacillus thuringiensis* to six forest Lepidoptera and *Bombyx mori*. *J. Invertebr. Path.* **62**, 295-301.

Van-Nguyen, A. (1995); Biochemistry and molecular genetics of a novel entomocidal toxin from *Bacillus thuringiensis*. PhD Thesis, University of Cambridge.

Van Rie, J., Jansens, S., Höfte, H., Degheele, D. & Van Mellaert, H. (1989); Specificity of *Bacillus thuringiensis* δ-endotoxins: Importance of specific receptors on the brush border membrane of the midgut of target insects. *Eur. J. Biochem.* **186**, 239-247.

Van Rie, J., Jansens, S., Höfte, H., Degheele, D. & Van Mellaert, H. (1990a); Receptors on the brush border membrane of the insect midgut as determinants of the specificity of *Bacillus thuringiensis* delta-endotoxins. *App. Environ. Microbiol.* **56**, 1378-1385.

Van Rie, J., McGaughey, W. H., Johnson, D. E., Barnett, B. D. & Van Mellaert, H. (1990b); Mechanism of insect resistance to the microbial insecticide *Bacillus thuringiensis*. *Science* **247**, 72-74.

Visick, J.E. & Whiteley, H.R. (1991); Effect of a 20-kilodalton protein from *Bacillus thuringiensis* subsp. *israelensis* on production of the CytA protein by *Escherichia coli*. *J. Bacteriol.* **173**, 1748-1756.

Von-Tersh, M.A., Slatin, S.L., Kulesza, C.A. & English, L.H. (1994); Membrane-permeabilizing activities of *Bacillus thuringiensis* coleopteran active toxin CryIIIB2 and CryIIIB2 domain I peptide. *Appl. Environ. Microbiol.* **60**, 3711-3717.

Waalwijk, C.A., Dullemans, M., van Workum, M.E.S. & Visser, B. (1985); Molecular cloning and the nucleotide sequence of the Mr 28,000 crystal protein gene of *Bacillus thuringiensis* subsp. *israelensis*. *Nucl. Acids Res.* **13**, 8207-8217.

Walters, F.S., Slatin, S.L. Kulesza, C.A. & English, L.H. (1993); Ion-channel activation of N-terminal fragments from CryIA(c) delta-endotoxin; *Biochem. Biophys. Res. Com.* **196**, 921-926.

Ward, E.S. & Ellar, D.J. (1984); Cloning and expression in *Escherichia coli* of the insecticidal δ-endotoxin gene of *Bacillus thuringiensis* var. *israelensis*. *FEBS Lett.* **175**, 377-382.

Ward, E.S. & Ellar, D.J. (1986); *Bacillus thuringiensis* var. *israelensis* δ-endotoxin: Nucleotide sequence and characterisation of the transcripts in *Bacillus thuringiensis* and *Escherichia coli*. *J. Mol. Biol.* **191**, 1-11.

Ward, E.S. & Ellar, D.J. (1987); Nucleotide sequence of a *Bacillus thuringiensis* var. *israelensis* gene encoding a 130 kDa delta-endotoxin. *Nucleic Acids Res.* **15**, 7195.

Warren, G.W., Koziel, M.G., Mullins, M.A., Nye, G.J., Desai, N., Carr, B. & Kostichka, N.K. (September 1994). World Intellectual Property Organisation Patent WO 94/21795.

Weinert, T.A., Schaus, N.A. & Grindley, N.D.F. (1983); Insertion Sequence duplication in transpositional recombination. *Science* **222**, 755-765.

Weiss, B. (1976); Endonuclease II of *Escherichia coli* is exonuclease III. *J. Biol. Chem.* **251**, 1896.

Whiteley, H.R. & Schnepf, H.E. (1986); The molecular biology of parasporal crystal body formation in *Bacillus thuringiensis*. *Ann. Rev. Microbiol.* **40**, 549-576.

Widner, W.R. & Whiteley, H.R. (1989); Two highly related insecticidal crystal proteins of *Bacillus thuringiensis* subsp. *kurstaki* possess different host range specificities. *J. Bacteriol.* **171**, 965-974.

Wolfersberger, M.G. (1990); Specificity and mode of action of *Bacillus thuringiensis* insecticidal crystal proteins toxic to lepidopteran larvae: Recent insights from studies utilising midgut brush border membrane vesicles. In *Fifth International Colloquium on Invertebrate Pathology and Microbial Control: 1990*, pp. 278-282; Society for Invertebrate Pathology, Adelaide, Australia.

Wolfersberger, M.G. (1991); Inhibition of potassium gradient driven phenylalanine uptake in larval lymantria dispar midgut by two *Bacillus thuringiensis* δ-endotoxins correlates with the activity of the toxins as gypsy-moth larvicides. *J. Exp. Biol.* **161**, 519-525.

Wolfersberger, M.G., Spaeth, D.D. & Dow, J.A.T. (1986); Permeability of the peritrophic membrane of tobacco hornworm larval midgut. *Am. Zool.* **26**, 74A.

Wong, H.C. & Chang, S. (1986); Identification of a positive regulator that stabilizes mRNAs in bacteria. *Proc. Natl. Acad. Sci. USA*. **83**, 3233-3237.

Wong, H.C., Schnepf, H.E. & Whiteley, H.R. (1983); Transcriptional and translational start sites for the *Bacillus thuringiensis* crystal protein gene. *J. Biol. Chem.* **258**, 1960-1967.

Wu, D. & Aronson, A.I. (1992); Localised mutagenesis defines regions of the *Bacillus thuringiensis* δ-endotoxin involved in toxicity and specificity. *J. Biol. Chem.* **267**, 2311-2317.

Wu, D., Cao, X.L., Bai, Y.Y. & Aronson, A.I. (1991); Sequence of an operon containing a novel δ-endotoxin gene from *Bacillus thuringiensis*. *FEMS Microbiol. Lett.* **81**, 31-36.

Wu, D. & Federici, B.A. (1995); A 20-kDa protein preserves cell viability and promotes CytA crystal formation during sporulation in *Bacillus thuringiensis*. *J. Bacteriol.* **175**, 5276-5280.

Wu, S.J. & Dean, D.H. (1996); Functional significance of loops in the receptor binding domain of *Bacillus thuringiensis* CryIIIA δ-endotoxin. *J. Mol. Biol.* **255**, 628-640.

Wu, D., Johnson, J.J. & Federici, B.A. (1994); Synergism of mosquitocidal toxicity between CytA and CryIVD proteins using inclusions produced from cloned genes of *Bacillus thuringiensis*. *Mol. Microbiol.* **13**, 965-972.

Yamamoto, T. & McLaughlin, R.E. (1981); Isolation of a protein from the parasporal crystal of *Bacillus thuringiensis* var. *kurstaki* toxic to the mosquito larva, *Aedes taeniorhynchus*. *Biochim. Biophys. Res. Commun.* **103**, 414-421.

Yang, Y.J. & Davies, D.M. (1971); Digestive enzymes in the excreta of *Aedes aegypti* larvae. *J. Insect Physiol.* **17**, 2119-2123.

Yaoi, K., Kadotani, T., Kuwana, H., Shinkawa, A., Takahashi, T., Iwahana, H. & Sato, R. (1997); Aminopeptidase N from *Bombyx mori* as a candidate for the receptor of *Bacillus thuringiensis* Cry1Aa toxin. *European J. Biochem.* **246**, 652-657.

Yap, W. H. & Porter, A. G. (1993); Enhanced expression of mosquitocidal toxin genes in recombinant *Caulobacter crescentus*. In *Society for Invertebrate Pathology XXVIth Annual Meeting*, Asheville, North Carolina, USA.

Yoshisue, H., Fukada, T., Yoshida, K-I., Sen, K., Kurosawa, S-I, Sakai, H & Komano, T. (1993a); Transcriptional regulation of *Bacillus thuringiensis* subsp. *israelensis* mosquito larvicidal crystal protein gene *cry*IVA. *J. Bacteriol.* **175**, 2750-2753.

Yoshisue, H., Nishimoto, T., Sakai, H. & Komano, T. (1993b); Identification of a promoter for the crystal protein-encoding gene *cry*IVB from *Bacillus thuringiensis* subsp. *israelensis*. *Gene* **137**, 247-251.

Yoshisue, H., Ihara, K., Nishimoto, T., Sakai, H. & Komano, T. (1995); Expression of the genes for insecticidal crystal proteins in *Bacillus thuringiensis*, *cry*IVA not *cry*IVB is transcribed by RNA polymerase containing σ^H and that containing σ^e. *FEMS Microbiol. Lett.* **127**, 65-72.

Yu, Y.M., Ohba, M. & Gill, S.S. (1991); Characterisation of mosquitocidal activity of *Bacillus thuringiensis* subsp. *fukuokaensis* crystal proteins. *Appl. Environ. Microbiol.* **57**, 1075-1081.

Yu, C.-G., Mullins, M. A., Warren, G. W., Koziel, M. G. & Estruch, J. J. (1997); The *Bacillus thuringiensis* vegetative protein Vip3A lyses midgut epithelium cells of susceptible insects. *Appl. Environ. Microbiol.* **63**, 532-536.

Yuan, K., Johnson, W.C., Tipper, D.J. & Setlow, P. (1981); Comparison of various low molecular-weight proteins from dormant spores of several *Bacillus* species. *J. Bacteriol.* **146**, 965-971.

Yudina, T.G. & Burtseva, L.I. (1997); Activity of delta-endotoxins of four *Bacillus thuringiensis* subspecies against prokaryotes. *Microbiology* **66**, 17-22.

Yudina, T.G., Yegrov, N.S., Loria, J.K. & Vybornykh, S.N. (1988); Biological activity of *Bacillus thuringiensis* parasporal crystals. *Izvestiya Akademii Nauk SSSR Seriya Biologicheskaya* **3**, 427-436.

Yudina, T.G., Mil'ko, E.S. & Egrov, N.S. (1996); Sensitivity of *Micrococcus luteus* dissociation varients to δ-endotoxins of *Bacillus thuringiensis*. *Microbiology* **65**, 321-325.

Yunnovitz, H (1986); A new sensitive method for determining the toxicity of a highly purified fraction from δ-endotoxin produced by *Bacillus thuringiensis* var. *entomocidus* on isolated larval midgut of *Spodoptera littoralis* (Lepidoptera, Noctuidae). *J. Invertebr. Pathol.* **48**, 223-231.

Zhang, J., Hodgman, C., Schnetter, W. & Schairer, H.U. (1996); Characterisation of a *cry* gene from *Bacillus popillae* subsp. *melolontha* H1; In *Society for Invertebrate Pathology 29th Annual Meeting:1996*, p. 93; Cordoba, Spain.

Zhong, C. (1996); Characterisation of a *Bacillus thuringiensis* δ-endotoxin toxic to three orders of insects. PhD Thesis, University of Cambridge.